

What kind of arms control?

The prospects of some success for the weary arms controllers at Geneva are brighter than for many years. The most obvious dangers are that there is almost too much choice.

NOT often does Geneva, where the arms control negotiators meet, become a source of cheerful rumour. But that is what it has been the past few weeks. People whose profession is cautious scepticism, habituated to the view that good things will be an uncovenanted bonus, are going about with a new spring in their step. But even those of us who are not privy to what the negotiators have said to each other have some grounds for optimism. First, the Geneva summit last year has made East-West relations more real than they used to be. Second, whatever may be the hidden traps in Mr Mikhail Gorbachev's new package of arms control proposals, announced at the beginning of this year, it is also full of opportunities that may be teased out. And, finally, this is an election year in the United States. After a few months of battling with the budget and the Gramm-Rudman Act, it will be a relief and an advantage to be able to hold up something done on arms control as a proof that this year has not been all dust and ashes. It is not mere cynicism to remark on President Reagan's wish that this year's instalment of the East-West summit should take place well before the election in November.

Two more substantial impediments to agreement seem also to have been implicitly, if only temporarily, removed. The Soviet Union has not abated its fierce opposition to the US Strategic Defense Initiative (SDI), but Soviet statements in the past few weeks have not so directly as in the past made the abandonment of SDI a precondition of any kind of agreement. To read between the lines, a deal on SDI may be a Soviet precondition for an agreement on strategic missiles, which naturally directs attention to lesser agreements on smaller but still important issues. The second development is what seems to be a loosening of Soviet views on the relevance of British and French nuclear weapons to a bilateral agreement on missiles of intermediate range and sited in Europe. To be sure, this flexibility may not mean much in the long run. The Soviet view a year ago was that the minor nuclear forces posed a strategic threat to the Soviet Union and must therefore be accounted for in an agreement on strategic arms; it is merely consistent that they should be left out of account in deciding what balance of intermediate-range weapons is fair, but without such an understanding, it is plain that nothing could be done quickly to make a European agreement stick.

Bigger prize

At the risk of counting unhatched chickens, it is therefore worth calculating the consequences of the biggest potential prize from the current Geneva talks, an agreement to dispense with missile-borne weapons in Europe. The rationale of such a deal, if for a time not linked with an agreement on strategic weapons, would be that the risk of nuclear war by accident would be reduced, that Moscow (like the rest of European Russia) would be less vulnerable to quick (ten-minute) attack, as would be Western Europe, and that the general climate would be improved. If these are the objectives, it makes sense to discount, if only for the time being, the damage that might be done by nuclear weapons carried by conventional aircraft. On the face of things,

the risk in an agreement along these lines would be that the defence of Western Europe, predicated as it is on the assumption that nuclear weapons will be needed to counter a conventional attack from the East, would be compromised by the removal of intermediate-range weapons. The truth is that, while the strategic forces remain, together with conventional forces, the circumstances would not be very different from those obtaining between fifteen and ten years ago. In retrospect, before the SS20 missiles were installed in Eastern Europe, that was a relatively cheerful time.

So there is no immediate reason why an agreement to remove Soviet and US missiles from within striking distance of the European theatre should not benefit both sides. That is the narrow calculation. More distantly, two more taxing issues would arise. First, the governments of Western Europe would find it logically incumbent on them to increase their spending on conventional forces, as the US Congress is forever crying that they should. But Europe, just now, is not in much of a mood to follow this course. Second, the mere abolition of missiles of intermediate range would not by itself create the flexibility in Central Europe which, to many Europeans, is the only benefit worth having from an agreement on European missiles. That, unfortunately, is the hard truth. The arms control process will have to go much further before people are again free to wander about without visas.

Limits

What else may be attainable? Optimism being what it is, the question cannot be suppressed. So there is an important truth about the process of arms control that negotiators, politicians and people like the rest of us should repeatedly recall: however urgent may be the need to make the world a safer place, there is a limit to the speed with which that can be safely done. The origins of that paradox are simple enough, and fully illustrated by the experience of agreements now extant, the Non-Proliferation Treaty for example, or the unratified SALT II agreement. Agreements with content, and which are not merely cosmetic, inevitably restrict the freedom of the signatories to behave as they would otherwise have done. But learning to live with restraints externally imposed is not an easy business; witness the heart-searching in Washington last summer when it seemed as if the commissioning of a new Trident submarine would breach SALT II. But there are also technical problems in giving substance to any treaty requiring verification, as the constant worries about the inspection of nuclear sites under the Non-Proliferation Treaty have shown. Yet the risks that would attend the collapse of a newly negotiated treaty are too great for comfort. That is why optimists should be hoping that progress will rather be steady than dramatic.

None of this implies that there is nothing else that can be done in the immediate future. To show willing, but also for the benefits that would accrue, it would make sense that the two superpowers at Geneva should agree to ratify the existing treaties on the Threshold Test-Ban (of 150,000 tonnes of TNT) and that regulating the use of peaceful nuclear explosions. In its



present mood, the US Congress would probably take these documents on the nod, while their administration would not impede other constructive work or be a technical nuisance. It would also be sensible that an agreement on intermediate missiles should be embedded in some of the constructive suggestions for making Central Europe a safer place that have emerged from the Stockholm negotiations.

So why not a comprehensive test ban? Because the United States has a military need to develop warheads for the missile that will have to replace the unworkable MX. (Making X-ray lasers with nuclear explosions must be less important.)

There is a psychological problem that cannot be lightly swept aside. For as long as there is no binding agreement on strategic nuclear arms, those who agree that there shall be no tests of nuclear weapons will know that the reliability of their stockpiled weapons, and their capacity to improve them, will be restricted, and that while the same restrictions may formally apply to the other fellow, there is always a chance that he will have got ahead. While defence establishments are run by specialists called soldiers, that anxiety is bound to sap conviction that a comprehensive test ban is the route to greater safety.

That must be the capstone on any process such as that outlined by Mr Gorbachev a few weeks ago. If all goes well this year, how soon can a start be made on that? Curiously, there seems already to be a measure of agreement on the stages by which an agreement must be reached. First, there would be a fifty per cent reduction by either side (which would not remove the need for the MX replacement). This, conveniently, would have the benefit of allowing the problem of the British and French nuclear weapons to be postponed (but not for long). The snag is that nothing can be expected of the strategic negotiations until there is an understanding about SDI, which is the rub. The best hope is for a compromise, a reaffirmation of the Anti-Ballistic Missile Treaty by the United States (because SDI has threatened to be in breach of it) and a renegotiation of the treaty to clear up its present ambiguities. Ideally, that agreement should precede an agreement on strategic arms, perhaps even by the summer. That, on present form, is the best to hope for. But it is also the least. □

What price freedom?

International archaeology has made a muddle. Better act now to avoid further trouble.

THE outcome of this year's prolonged squabble about the attendance of scientists from South Africa at the planned Congress of the International Union of Prehistoric and Protohistoric Sciences (IUPPS) is about as bad as it could be. The Southampton organizers, despite the withdrawal of the *imprimatur* of IUPPS and the resignation of most of the independent members of the British organizing committee, have decided to go ahead with a meeting of some kind in the summer. Whatever other considerations may have weighed with them, they are probably conscious of how much effort and even money might have gone to waste if they had bowed to pressure.

Meanwhile, IUPPS itself plans a quite separate congress at Mainz in West Germany, although the details have not yet been settled. It is not, for example, yet known whether the alternative meeting will be timed so as to coincide with that at Southampton or, more discreetly, to be at a quite different time, perhaps not until next year. Inevitably, neither of these occasions will be as successful, either as an intellectual occasion or as a commercial enterprise, as it would otherwise have been. Worse than that, there is every chance that the animosities that have accumulated during the past few months will persist for years to come; principle and pride have conspired to drive a wedge between groups of people whose memories are inconveniently durable at the best of times.

With all this milk already spilled, the best hope now is that it

will be possible to avoid similar troubles in the future. The scandal at Southampton, it will be recalled, is that the plan to hold a conference open to all individuals wishing to attend was subverted by the views of people on the spot that only a boycott of people from South Africa would properly express their disapproval of South African policies of apartheid. The organizers began well enough, by insisting that those attending the meeting would do so as individuals, not as members of national delegations, which is how international meetings should be organized. Time will no doubt tell how vigorously they protested when the clouds of protest began to gather. But there is no doubt of what their course should have been; they should have cancelled the meeting once it became apparent that it could not be held freely.

Two sets of issues need attention, one of which is domestic. The groups at Southampton threatening to cloud the planned congress included the local students' union, the Association of University Teachers (AUT), the local branch of the Anti-Apartheid Movement and the City Council, which had earlier promised assistance with accommodation and hospitality. Among these groups, the big surprise is AUT, an organization which represents teachers at British universities in salary negotiations, but which has accreted a policy of "economic and academic" boycott of South Africa over a period reaching back to the late 1970s. At the outset, when resolutions passed at annual conferences were relatively innocuous, it must have seemed to the participants that a show of disfavour for the practices of the South African government would do no harm. Somewhere along the line, the union's policy has been hardened in a way that is bound, in the long run, to damage the professional interests of the majority of its members — their common interest that the practice of scholarship should be free. Although the local branch of AUT did its best, when the crisis at Southampton had burst into the open, to find a compromise acceptable to all, the mere fact of the union's declared policy proved an insurmountable obstacle at that late stage. But so long as the policy remains, there is no hope that international conferences can be held again in Britain without provoking similar trouble. It is therefore high time for the ordinary members of AUT to try to change their union's reckless policy on this important issue.

Internationally, the situation is even more confused. The statutes of IUPPS require that its conference should be open to all, but IUPPS is affiliated to a UNESCO umbrella organization which advocates an academic boycott of South Africa. Can that make sense? And what is to be made of the claim by the Southampton organizers that it is no longer possible to persuade scientists from countries whose governments are properly openly hostile to South Africa because of its apartheid policies to attend international conferences at which South Africans are present as individuals? Over the Southampton congress, the Anti-Apartheid Movement claimed to have written to at least eight such governments, ostensibly to alert them to the possible presence of individual South Africans. This intelligence seems to have had a powerful effect at Southampton, even though there is no evidence that the governments addressed would have been able and willing to prevent the attendance of those among their citizens wishing to attend. But if it is true that "third world" governments or, worse, scientists working there, will no longer allow themselves to be seen in the same room as colleagues working in South Africa, there is an urgent job for the scientific community to take on. Somehow, people have to be persuaded that there is a crucial distinction, in equity as well as common sense, between people of a particular nationality and the policies of their governments. Similarly, there is a distinction that must be drawn between the contributions that people can make, as scientists, to international conferences and the views that they may hold about the policies of their governments. That the Southampton organizers have compromised their case by banning some of the South African government's most courageous indigenous critics is in principle irrelevant, but serves to emphasize what a mess they have made of things. □

Challenger catastrophe

Commission concentrates on suspect rubber sealing rings

Washington

THE independent commission appointed by President Reagan to investigate the explosion of the space shuttle Challenger has concluded that the decision to launch the shuttle on 28 January "may have been flawed", it was announced last week. The 12-man commission, chaired by William P. Rogers, a former Secretary of State, has asked the National Aeronautics and Space Administration (NASA) to exclude officials concerned in the launch decision from NASA's own study of the disaster.

The announcement came at the end of a week in which it had become increasingly clear that the suspected cause of the accident, failure of the seals (known as O-rings) on the right solid rocket booster early in the flight, had been a major concern of NASA safety engineers for up to three years.

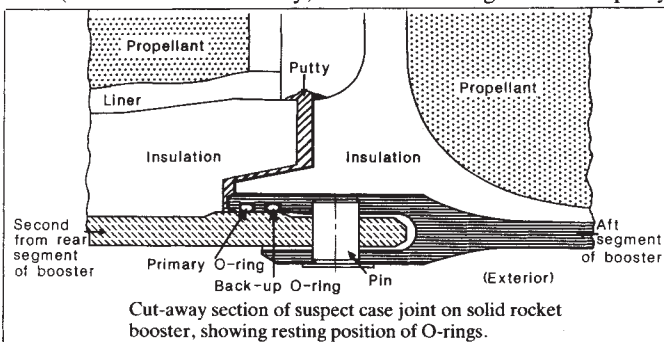
Internal documents released by NASA show that damage to the rings had been observed in several previous flights, and by last summer it had become a priority issue. A memorandum written by a junior NASA budget analyst dated 23 July 1985 records that management was "viewing the situation with the utmost seriousness". A new design of booster now awaiting testing incorporates a special safety feature to ensure that back-up O-rings work at all times.

Although the cause of the accident has still not been formally decided, a number of clues point to a failure of O-ring seals on the aft case-case joint of the right solid rocket booster as the culprit. A public meeting of the commission last week in Washington was devoted to the possibility.

The case-case joints link the four segments that make up the body of each booster. A seal failure at one of the joints would be expected to lead to a breach in the casing of the rocket, which could then in turn have caused the explosion of the external fuel tank, 74 seconds after launch. An unusual plume of dark smoke is visible near the suspect joint in engineering photographs taken as little as 0.4 seconds after launch; 60 seconds later, a jet of flame is seen emerging at about the same point. Telemetry data indicate that the right booster lost up to 5 per cent of its thrust during the brief flight, consistent with a breach of the casing.

More disturbing is the evidence, first made public last week by the *New York Times*, that NASA had long been aware of potentially serious problems with the O-rings. The two rings, a primary and a back-up at each joint, are designed to provide a seal to withstand the rapid rise of pressure in the boosters upon firing, from zero to 900 pounds per square inch in about 0.6 seconds. Protecting the O-rings from the hot exhaust gases is a layer of zinc chromate putty inserted into the gap between the insulation on adjacent segments.

It now emerges that NASA has known since at least December 1982 that the rings do not "seat" and provide a good seal instantly, so "blow-throughs" of the putty



Cut-away section of suspect case joint on solid rocket booster, showing resting position of O-rings.

occur, allowing the hot gases to rush past and erode the primary O-rings. Six of 171 case-case joints examined have shown significant erosion of the primary O-ring, leading in two cases to deposits of soot behind the ring. Joints between case and nozzle have an even worse record, with 16 instances of primary ring erosion out of 57 joints examined.

The problem is more critical in case-case joints because bending of the joints due to the pressure build-up unseats the back-up O-ring shortly after ignition. As a result there is a "high probability of no secondary seal capability" after 330 milliseconds post-ignition, according to documents prepared by the solid booster manufacturer, Morton Thiokol Inc. NASA recognized this early in 1983 and reclassified the primary O-rings as "single failure points" where one failure would mean loss of the orbiter and the crew.

Particular attention has focused on the possible role of the cold weather at the launch site on 28 January. The temperature, 38 degrees Fahrenheit, was 13 degrees colder than at the previous coldest launch; it reached a minimum of 24 degrees during the previous night. NASA officials testifying last week assured the commission that the decision to launch was sound because the most erosion ever

Feynman's physics

Washington

LAST week's meeting of the presidential commission on the shuttle disaster was a solemn occasion, but not enough to prevent Richard Feynman, physicist and practical joker of renown, from surprising onlookers and fellow commission members with an impromptu physics experiment that may turn out to get to the heart of the Challenger disaster.

Commission members had been invited to pass around and examine a life-sized segment of the suspect case joint, complete with sections of O-rings. Unannounced, Feynman extracted one of the O-ring sections and disappeared to get some iced water. Later he was seen borrowing a small G-clamp from a microphone stand, and then conducting some experimental manipulations on the table in front of him. After asking a NASA official some pointed questions about how the resilience of the O-rings was affected by temperature and securing an admission that this was crucial for integrity of the joint, Feynman held up the borrowed O-ring segment, which he had compressed to about half its diameter with the G-clamp and suspended in iced water. He then released the G-clamp and pointed out that the cold had made the rubber O-ring lose its resilience; it took about a minute to return to its proper shape. "I believe", said Feynman, "that this has some significance for our problem".

Tim Beardsley

observed in a case-case joint had been at most one-third of that likely to cause failure. But responding to questioning from commission member and physicist Richard Feynman, Lawrence Mulloy, NASA's solid rocket booster project manager, agreed that low ambient temperatures could have caused the synthetic rubber rings to become less resilient; nevertheless, the rubber was supposed to retain its elastic properties to -40 degrees.

Mulloy told the commission that concern about the effect of cold on the boosters was expressed the day before the launch by Morton Thiokol engineers, who initially recommended that the launch be postponed if the ambient temperature was lower than the previous lowest temperature launch (51 degrees Fahrenheit). But after a full discussion with NASA, it was decided to go ahead because there was "no direct correlation" between O-ring erosion and ambient temperature.

NASA has often been criticized for the frequent delays to shuttle launches in the past, and the perceived pressures not to delay further must have been considerable. The President's commission is now said to be concerned that apprehensions on the part of those responsible for determining flight readiness were not given appropriate attention.

Tim Beardsley

Monkey business

Bolivia asks for animals back

Washington

THE government of Bolivia is demanding the return of 361 owl and squirrel monkeys exported to the United States last month for use in the development of a malaria vaccine. The animals, the first of a lot of 600, were exported under a special ministerial exemption to Bolivia's existing ban on the export of fauna and are now under quarantine in Miami. Last week the ministerial exemption was annulled and the rest of the shipment is unlikely to leave Bolivia.

The animals were imported by Worldwide Primates Inc. of Miami, run by animal dealer millionaire Matthew Block, and were to have been distributed to more than 14 US institutions supporting malaria research. The animals are not considered seriously at risk of extinction, although it is argued that there has been no adequate census of the animals and that it would be possible to use captive-bred animals for the research.

The affair led to frank disagreement between the State Department, which has supported Bolivia's efforts to reduce illegal animal exports, and the US Agency for International Development (AID), which coordinates malaria research. Two years ago, Bolivia banned the export of all fauna (although enforcement is generally considered lax) and so the State Department declared its opposition to the deal. Bolivia is also a signatory of CITES, the Convention on International Trade in Endangered Species.

Squirrel monkeys and owl monkeys (*Saimiri sciourens* and *Actus trivirgatus*) are listed in Appendix 2 of CITES, which means they are considered threatened but not "endangered". US laws permit the import of Appendix 2 animals with proper documentation.

The proposal to export the animals to the United States was apparently made by Bolivian scientists at a conference on malaria in Florida last autumn, and the Bolivian government agreed. State Department officials were concerned about the setting of a precedent, but backed down when it became clear that Worldwide had secured a government exemption. Dr Kenneth Bart of AID's health office later criticized the State Department for "inappropriately" describing the species as "endangered".

There has been a worldwide resurgence of malaria in the past few years, primarily because of the drug resistance of malaria parasites and insecticide resistance in the mosquito vectors. The disease causes 2-3 million deaths each year, mainly among young children. There has been significant progress towards vaccines in recent years thanks to genetic engineering, but the dif-

ficulty of obtaining non-human primates for testing (required by the Food and Drug Administration before human volunteer trials can begin) has become a major bottleneck. AID has supported a network of laboratories working towards a vaccine for 15 years.

The Bolivian Embassy in Washington will offer no explanation for the exemption or its withdrawal; AID officials blame hysteria in the Bolivian press and political machinations by opponents of the government.

The details of how the ministerial resolution allowing the export (no. 347/85) was obtained are intriguing. The resolution names as Worldwide Primates' representative in Bolivia a company called Altros Internacional, although the CITES export certification names Aves Import as the exporter. The animals already in the United States (341 squirrel monkeys and 20 owl monkeys) have a declared value of \$55,150. The resolution reversing the earlier exemption is no. 45/86; it is unclear

whether the animals (now the property of AID) will be returned to Bolivia to be released to their native habitat, as the government is demanding.

Officials in the Bolivian department of agriculture were initially introduced to close family relatives of Altros's owner by Gene Harris, a US travel agent; a US AID official certified that the animals were to be used for *bona fide* medical research and would later go to breeding colonies. Later, following highly critical coverage of the affair in the Bolivian newspaper *Ultima Hora* and (ungrounded) suggestions that the animals were destined for pet shops, Altros decided to have nothing more to do with the deal, and Aves solicited and received the CITES export document. Altros then brought a civil suit against Block and is said to be seeking compensation for having its good name besmirched; however, an arrest warrant was put out for Block, and his passport was confiscated. Block has nevertheless managed to return to the United States. The US State Department is uncommunicative, saying the United States takes no formal position on either the exemption or its withdrawal. **Tim Beardsley**

AIDS

Japan screens donated blood

Tokyo

JAPAN is waking up to the fact that it is not immune from acquired immune deficiency syndrome (AIDS). Japan's Health and Welfare Ministry has rushed to implement a programme to screen a large proportion of the nation's blood donors, and an emergency grant has been awarded to develop new diagnostic tests for AIDS. These developments come close on the heels of approval of the import of an AIDS antibody test kit manufactured by Abbot Laboratories of the United States and confirmation of three more cases of AIDS in Japan, bringing the total to fourteen.

Japan's Red Cross will test the blood of one million donors in Tokyo and Osaka over a period of one year from 17 February, thereby screening more than 10 per cent of the 8 million expected donations. As recently as November, the Health and Welfare Ministry had stated that mandatory screening of blood donors for AIDS was unnecessary and that the estimated cost, ¥10,000 million (about £38 million), was prohibitively high (see *Nature* 318, 306; 1985), but increased media coverage made it more urgent to allay public concern.

Pamphlets have been distributed by the Health and Welfare Ministry reassuring the public that if they live a "normal" life they have almost no chance of catching AIDS, and prefectural and municipal health officials were recently given a semi-

nar on the disease. But ignorance is still rife, and the conviction lingers that AIDS is somehow a "foreign" disease.

A dial-for-AIDS-advice service in Tokyo, which earlier last year was receiving only a few calls per month, is now inundated with questions centred on the fear that contact with foreigners may cause infection. The report that one of the latest AIDS sufferers, a foreigner, associated with Japanese men because he thought they were free of AIDS will no doubt only add fuel to such suspicions.

The Red Cross will use an enzyme-linked immunosorbent assay (ELISA) antibody test kit supplied by Dynabot, the Japanese subsidiary of Abbot Laboratories, for the block screening programme. It was announced last week, however, that an emergency grant of ¥40 million has been awarded by Japan's Science and Technology Agency to two research teams at the Health and Welfare Ministry's Health and Medical Affairs Bureau and the National Institute of Health to develop more reliable AIDS tests.

In addition to the testing of one million donors, ¥50 million has been put aside this year by the Japanese government for blood screening, but the funds allocated for AIDS research in Japan pall to insignificance compared with the more than \$300 million (about ¥55,000 million) expected to be spent by the US Public Health Service on AIDS research this year. **David Swinbanks**

Pharmaceuticals

NIH retreat from controversy

Washington

US RESEARCHERS are querying the cancellation by the National Institutes of Health (NIH) of a major international conference on the use of antibiotics and antibiotic resistance. Plans for the conference were laid five years ago, but pharmaceutical industry executives accused the organizers of spreading "propaganda" and of "obvious pre-judgement" of the issues. Plans for the conference have been repeatedly changed and delayed, and NIH director James Wyngaarden has now decided to substitute a small workshop of conference planners.

The conference, administered by NIH's Fogarty International Center, was finally to have taken place next month. Wyngaarden says his decision to convene instead a small workshop was made exclusively on the basis of a scientific review of conference planning documents by acknowledged experts. But Dr Jerry Avorn of Harvard University, one of the organizers, said the conference documents are "every bit as good as the prevalent high standards of NIH" and that he was "left wondering" if lack of scientific merit was the true reason for the cancellation.

After a preparatory meeting for the conference in September 1984 at which six task forces were organized to prepare working documents, several industry executives wrote to Edward Brandt Jr, then acting assistant secretary for health, protesting that the organizers were biased against industry. The conference would have covered misprescribing and overuse of antibiotics both in humans and in animal feed. M. Marion Jones, president of Beecham Laboratories, said that the conference organizers were simply trying to confirm the beliefs of a number of "activists" present and to secure endorsements of proposals by NIH and the World Health Organization, which was also involved in planning.

Among those present at the meeting were Dr Sidney Wolfe, director of the Public Citizen Health Research Group, Dr Andrew Herxheimer of Charing Cross Hospital (London) and editor of *Drugs and Therapeutics Bulletin*, published by the British Consumers' Association, and Ms D. Melrose of Oxfam, all of whom have expressed concern about misprescribing of antibiotics.

Thomas Christie, vice-president of medical affairs of Wyeth International Ltd, wrote of a "general anti-industry bias" and said that further meetings of one of the task forces would "culminate in an outright attack on free enterprise". Christie asked Brandt to "use the influence of your office to limit this activity".

Wyngaarden says he was alerted to con-

cern about the conference by several individual members of conference planning groups; one, for example, wrote that an international conference would be premature because there are insufficient data. He then had several of the planning documents reviewed by experts (including himself) and four reviewers agreed the conference should be scaled down, although two recommended going ahead. Hopes for a large conference have not been entirely abandoned, however; the workshop next month may lay the groundwork for a large conference later.

Tim Beardsley

Shcharanskii release triggers hopes

Stockholm

THE release from the Soviet Union last week of Anatolii Shcharanskii after nine years in prison for alleged espionage appears to have raised spirits not only in the West but also within the Soviet Union. In Moscow last Saturday, a group of refusniks at a special meeting to mark the 850th anniversary of the birth of the mediaeval scholar Moshe Ben Maimon (Maimonides) issued a statement hoping that "those who have been striving for the simple and natural right to live in Israel will soon be free to do so".

The Maimonides seminar was the biggest gathering of refusniks since 1980. More than 40 people took part, including two visitors from Scandinavia, Dr Jens Larsen of Denmark (who spoke of last year's Niels Bohr celebrations) and Dr Oyvind Gron of Oslo (who spoke on the expanding Universe).

Apart from abstracts of recent work presented by almost 30 refusniks, there was also a poster session which appears to have been stimulated by the efforts made in recent years by scientists in the West who have been seeking means by which refusniks might contribute to international conferences by means of posters displayed *in absentia*.

Ironically, the Moscow seminar almost coincided with a meeting held in London by the Institute of Physics to mark the ninth anniversary of the arrest of Dr Yurii Orlov, now exiled to Yakytria, Siberia. Since the expiry of Orlov's five-year prison sentence, there have been several campaigns to persuade the Soviet authorities to commute internal to external exile.

Last week, the British lawyer John MacDonald announced that he had now written to the Soviet authorities to argue that, since Orlov had now turned sixty (pensionable age in the Soviet Union), he should be allowed to settle in Britain. Vera Rich

IVF

India embraces test-tubes

New Delhi

ALTHOUGH population control remains India's chief objective in human reproduction technology, *in vitro* fertilization (IVF) is now also officially encouraged as a treatment for infertility. With the first test-tube baby on the way at the Kem hospital in Bombay, more than a hundred infertile couples are being treated at the private clinics springing up in Bombay and Madras as well as at the government-run All-India Institute of Medical Sciences, which has set up an IVF clinic here.

"With so many of our women wanting to go abroad for test-tube babies, we cannot ignore the need for treatment in India", says Dr Kamal Bukshee, the gynaecologist in charge of the All-India clinic. Although India may be oppressed by too many children, physicians are concerned that infertility, which is a significant problem in India, may be exploited by quack sexologists.

One of the British hospitals offering IVF treatment has recently appointed an agent in New Delhi to handle enquiries from Indian patients seeking treatment. According to Dr Bukshee, many of those enquiring are prevented from going abroad by the prohibitive cost. She has herself developed a variation of the standard technique in which ultrasound is used for scanning the ovary and locating ripe eggs, which are then removed by the use of a needle rather than by laparoscopy. She says that treatment can be offered in the outpatients' department of the hospital at a fraction of the cost that patients would have to spend if they went abroad.

In vitro fertilization was first sanctioned three years ago by the Indian Council for Medical Research, which after much hesitation began a programme of research at its Institute for Research in Reproduction (IRR) at Bombay. In collaboration with the Kem hospital, the institute has now perfected the method after trying it out with 25 women with blocked fallopian tubes. The first success was with a 30-year-old janitor who is now four months pregnant. This has so stimulated demand that a computer is now being used to store information about the women on the waiting list, and to choose which should have treatment.

Dr T.C. Amand Kumar, director of IRR, says that even if all the women now seeking treatment were to become pregnant, "there will be no undesirable effect on our population", but that IVF technology will have beneficial effects in medicine as a whole, especially in the treatment of inherited diseases by gene therapy.

K.S. Jayaraman

US plant science

Monsanto's shock firings

Washington

PLANT science laboratories across the United States have in the past month been deluged with the *curricula vitae* of the 80 or so researchers recently "deployed" from Monsanto Corporation's plant sciences group in St Louis, Missouri. Monsanto claims that the cuts reflect the deliberate and sensible shutdown of several unfeasible projects; even so, its action has roused a sense of foreboding among the many laboratories, both academic and commercial, that viewed Monsanto's exemplary programme as a prototype.

The chemical company's reputation for long-term commitment to basic research has lured top talents to its benches for decades. Now, in the wake of a 1 January consolidation that erased the plant sciences group as a separate entity, a team that was once more than 150 strong has probably dwindled to fewer than 50. Monsanto has eliminated its plant growth regulator and soybean hybridization projects, which seemed unlikely to yield returns in less than 15 years, but has retained traditional herbicide and fungicide programmes in the newly-formed Monsanto Agricultural Company. The basic research efforts that escaped the scythe have been subsumed under Monsanto's central research division, taking several reassigned plant sciences researchers with them.

The first wave of deployments struck at the end of October, at about the time Monsanto completed the \$559 million purchase of G.D. Searle. Twenty employees were informed that they were "relieved of their responsibilities", and would be paid until 31 March 1986, during which time Monsanto would either find them other posts within the corporate research structure, or help them to find jobs elsewhere.

Monsanto contends that the deployments were part of a larger, pre-announced restructuring effort, but they came as "quite a shock" to the plant sciences workers, especially as some of those concerned had joined Monsanto just two months previously. (Monsanto says these hirings resulted from "miscommunications" between corporate executives and the personnel department.) Sixty more researchers were given notice on 20 December.

Critics outside the company suspect that Monsanto's manoeuvres may not be the result of sagacious planning. Restructuring of the company caused a \$98 million net loss in 1985, compared with a net profit of \$439 million in 1984. And Monsanto's agricultural interests, in particular, may be threatened soon when the glyphosate and alachlor herbicides which generate over 80 per cent of agrichemical profits

go out of patent. Small wonder, then, that the fledgling agricultural company has turned its attention to herbicide and fungicide research, whose rewards may accrue more quickly.

Plant researchers are worried, however, that other companies may share Monsanto's pessimism, if not its plight. They fear that industry may launch long-term projects in agricultural genetics with expectations of short-term profits, then abandon these projects when the gap between expectations and returns begins to open up. Although genetic manipulation is a reality, plant physiology and pathogen

interaction are in many cases too poorly understood to show which genes should be manipulated. Product development in these areas will probably take at least ten years, by conservative estimates.

Some companies seem to have acknowledged these delays, keeping long-term plant sciences projects scaled down and short-term projects busy. Dow and DuPont, for example, maintain active plant genetics programmes, but count on their herbicide divisions for immediate returns: DuPont is anticipating an enthusiastic reception for its sulphonylurea herbicides, and Dow registered its latest herbicide, tridiphane, two weeks ago. When the dust settles, it may be that Monsanto's new structure will simply conform to this pattern.

Karen Wright

Indian science

All change at the top

New Delhi

AFTER liberalizing the import of overseas technology and giving a new direction to the Indian government's support of science (see *Nature* 314, 396; 1985), Mr Rajiv Gandhi, the Prime Minister, has now decreed a radical reshuffling of the senior science bureaucrats.

The chief casualty is Dr T.N. Khoshoo, secretary of the Department of the Environment, who was urgently recruited from his previous post at the Indian Institute of Botany at Lucknow just over two years ago. Dr Khoshoo will leave the bureaucracy. He has been replaced by a non-scientist.

During the past few days, it has also been announced that Dr Yash Pal, previously secretary of the Department of Science and Technology (DST), which plays a central coordinating role as well as being one of the chief sources of research grants for university people, is to be moved to the University Grants Commission. Pal's immediate predecessor at DST, Dr S. Varadarajan, who has been secretary of the Department of Scientific and Industrial Research (DSIR) for just over a year, and who in that role has been responsible for more than 40 national research laboratories, has been transferred to the Planning Commission, not as a full member but as its chief consultant, a post that carries no well-defined responsibilities.

Mr Gandhi has also made a radical change in the machinery by which the Indian government receives advice on science. The Scientific Advisory Committee to the Cabinet, consisting of some 22 top science bureaucrats, and set up only in 1982, has been abolished. Instead, there will be a new Scientific Advisory Committee to the Prime Minister consisting of seven people drawn from non-government laboratories and from private

industry. The chairman of the committee is Dr C.N.R. Rao, director of the Indian Institute of Science at Bangalore and president of the Indian National Science Academy.

Although the changes are ostensibly part of the Prime Minister's scheme to clear out dead wood and to improve the performance of the science establishment, they have inevitably provoked confusion and uncertainty. It is not clear whether the government means business or is, instead, using the changes as a means of shock therapy. One particular confusion is that the posts vacated by Pal and Varadarajan are for the time being left empty.

The removal of Varadarajan has been a big surprise. He was specially chosen as head of DSIR by the late Prime Minister, Mrs Indira Gandhi, because of his varied experience in the chemical industry as well as at government laboratories.

Another surprise is the appointment of Professor M.G.K. Menon to the new post of Science Adviser to the Prime Minister. Only a few days earlier, Menon had accepted a five-year appointment as research professor from the Indian National Academy of Science. He is known to have been unhappy with the government's policy on technology imports, which contradicts the recommendations of the 1982 policy statement that he helped to frame.

Critics of the changes now announced also say they are doubtful of the wisdom of replacing the old advisory council with the smaller group now chosen, which will have the responsibility of advising the Prime Minister on the major issues that will arise between now and the end of the century. How, the critics are asking, can three research chemists, a neurosurgeon, a theoretical physicist and an aeronautical engineer tackle the task laid down for them, to provide a perspective plan for the year 2001?

K.S. Jayaraman

Japanese broadcasting satellites

Toshiba pays price of failure

Tokyo

THE successful launch last Wednesday of Yuri-2b, Japan's second direct television broadcasting satellite, has not quietened critics of the project. A series of delays and failures had slowed progress and raised costs. One casualty had been the Toshiba Corporation, the main contractor. In an unprecedented move, it has been dropped from the project and replaced by NEC.

Another casualty might be said to be the general public, because the satellites are intended for the national broadcasting service, NHK. One newspaper urges its readers to ask how much money has been poured into the project next time the man from NHK comes around to collect payment of the television licence.

Problems really began to attract notice when the first broadcasting satellite, Yuri-2a, broke down soon after it went into operation in May 1984, leaving only one of its three transponders functioning (see *Nature* 309, 295; 1984). The fault seems to have been in travelling wave tubes designed to beam television signals to the Japanese islands and supplied by a French and a US company. These failures provoked furious complaints to the visiting US Vice-President George Bush about the quality of foreign workmanship. And to take the blame the president of the National Space Development Agency (NASDA) was forced to resign.

Next, it was decided that Toshiba would be replaced as prime contractor by NEC for future broadcasting satellites. This is a serious blow to Toshiba and may put it out of the satellite business altogether. Japan's three long-term satellite development programmes, in communications, broadcasting and weather observation, had been given out one each to Mitsubishi Electric, Toshiba and NEC so that each can independently acquire sufficient expertise eventually to compete with each other and with foreign companies. NEC will now continue the broadcasting series. Next in line are the BS-3 satellites which will be launched at the end of the decade. They are intended to broadcast high-definition television (HDTV) signals which will give a picture quality as good as 35-mm film projection.

It may be that much of the high cost and the failure of the broadcasting satellite programme is due to the particular development strategy Japan is pursuing. Satellites can be bought much more cheaply abroad than they can be built in Japan, even though only a third of the components of the Yuri satellites are Japanese. But satellite imports have been banned in order to allow domestic industries to master construction technology.

The rush to build broadcasting satellites — Yuri-2a was the first in the world — is intended to provide a lead in direct broadcasting technology. In particular, Japanese companies are looking at the potentially huge market for the roof-top parabolic antennas and decoders needed to pick up satellite transmissions.

In order to make these small and easily marketable, the Yuri-2a satellite's transmitters were much more powerful than had previously been tested in space. It was these that broke down. The just-launched Yuri-2b satellite has the same equipment aboard but all has been checked and re-checked, putting back the launch date by six months. That in turn has delayed other launches and the testing of the new H1 rocket with its home-developed liquid

oxygen/hydrogen engine.

Delays have also lessened the chances of international acceptance of the HDTV standards developed in Japan. It was hoped that speedy development of the BS-3 series would enable NHK to introduce high-definition broadcasts before anyone else, and to seize the initiative in setting international standards. The present satellite will be used for experiments with such television broadcasts and super high-fidelity FM digital broadcasts. Its main purpose however, will be to eliminate poor reception in outlying and mountainous regions of Japan.

Whether lost time can now be made up depends on the success of the satellite when it is switched on in a couple of months. The satellite's insurers are clearly not totally confident — the premium has been increased to 30 per cent of the satellite's value from the 12 per cent set for the previous satellite.

Alun Anderson

Malaria vaccines

Biogen revives forgotten project

COMMERCIAL interest in the production of a malaria vaccine has taken another turn with the decision of Behringwerke AG, the subsidiary of Hoechst AG based in Marburg, West Germany, to take over a project started by Biogen, the Boston and Geneva-based biotechnology company, which has been through a bad patch.

Indeed, the Biogen malaria vaccine project was a casualty of the company's problems over a year ago which led to a 20 per cent cutback in staff, and the removal of both Professor Walter Gilbert as chief executive officer and Dr Julian Davies as research director in Geneva.

For the past year or so, malaria vaccine work at Biogen has continued only because members of staff are allowed to spend 20 per cent of their time on their own projects and because of collaborations with Dr Luc Perrin at the Geneva Blood Centre and Dr Bernard Mach at the University of Geneva.

The result, say observers, is that the Biogen project has fallen behind its rivals and that the company has been lucky to strike a deal with Behringwerke.

The Biogen approach to a vaccine has concentrated on the blood-borne stage (merozoite) of the *Plasmodium falciparum* parasite, and particularly on a large protein that is exposed on the merozoite's surface. Wellcome Biotechnology in Britain and Hoffman-La Roche in Basel, in collaboration with Dr John Scaife's group at the University of Edinburgh, are pursuing the same approach, but seem to be considerably closer to a real test of whether the protein will protect experimental monkeys against malaria.

Other merozoite proteins are being investigated as vaccines at the Walter and

Eliza Hall Institute of Medical Research in Melbourne, Australia, and in Sweden at the University of Stockholm. And vaccines based on proteins on the surface of the sporozoite stage of the parasite, the form in which it is passed from the mosquito into the blood stream, are in an advanced stage of development both at New York University, probably in conjunction with Hoffman-La Roche, and at the Walter Reed Army Institute of Research in Washington, DC, in conjunction with Smith Kline and French Laboratories.

It is no coincidence that the military are involved in the last project, for it is military personnel who form one of the prime targets for vaccination, at least where money is no object. A second group in the same category are travellers from developed countries to areas where malaria is endemic. Figures that emerged in London last week show the steady rise in cases of malaria among that group. Over 2,000 cases of malaria among travellers from Britain were recorded last year and 85–90 per cent of the cases were the result of failure to take medication as instructed. If ever there is an effective vaccine, that problem would be solved.

While Behringwerke tries to make the most of Biogen's malaria work, Biogen is developing a new strategy under its chief executive of two months, James Vincent. His first task is to improve the financial position of the company, which took a considerable turn for the worse to judge by the year-end figures released last week. Total revenues for 1985 at \$21 million were \$10 million down on 1984, giving a net loss in 1985 of \$19 million compared with \$13 million in 1984. **Peter Newmark**

French science policy

Good effort, but must try harder

THE French government has made "undoubted progress" and achieved "significant results" with its innovation policy, but "serious weaknesses" remain. These are some of the conclusions of a report requested a year ago by the French research minister Hubert Curien from the Paris-based Organisation for Economic Cooperation and Development (OECD), the international organization whose members include 24 of the world's richest nations. OECD has a long record of making assessments of national science policy.

OECD is determinedly politically independent, but its report on France, the third of a series of national innovation policy reviews, but the first to deal with a major economy, comes at a critical juncture, as it appears less than a month before the French general election on 16 March, which is expected to throw out the present socialist government. Moreover, research policy is probably that government's greatest success, and the report is therefore being treated with great political seriousness as a clear objective analysis of the government's past and present programme, as developed under socialist president François Mitterrand, who came to power just five years ago.

Despite difficult economic circumstances the world over, French researchers have not lacked money. The French government has put an immense effort into research and development compared with its economic competitors, according to figures in the OECD report.

Comparing 1983 levels with those of 1979, French spending on basic research is estimated to have risen 43 per cent in real terms, compared with 9 per cent in the United Kingdom, 6 per cent in Japan, 1 per cent in the United States and no rise at all in West Germany. Government support for industrial research and development rose 48 per cent in France over the same period, followed by West Germany (22 per cent) and Japan (10 per cent). Over the same period, support for industrial research actually fell in the United Kingdom (by 4 per cent) and in the United States (by 6 per cent).

France was also ahead in every other category of research spending, according to OECD, such as infrastructure (a real increase of 18 per cent over the four years, the nearest rival being Japan with 16 per cent), health and welfare (33 per cent, compared with falls or stagnation in other major countries) and space and defence research (26 per cent, the nearest rival being the United States with a 21 per cent increase).

But despite this cash flow, the fundamental conclusion of the authors of the

report, a group of senior, non-French European and US technology policy-makers, is that France still has a very long way to go before it can become as competitive in innovation as West Germany, the United States, Japan or even Great Britain.

During the whole of the 1960s and 1970s for example, the proportion of patents granted in France to companies of French origin fell by nearly half, from 40 per cent in 1960 to 25 per cent in 1980. Even if President Mitterrand's policies manage to reverse this trend, the effect will be long in coming.

According to a poll conducted by the European Management Forum, a group representing senior management in Europe, France is actually perceived to have dropped in its capacity to innovate since 1980, falling from ninth among nations in 1980 to thirteenth in 1985, using criteria such as "capacity to exploit inventions commercially" and "capacity to abandon declining activities and invest in growth activities".

Nevertheless, OECD broadly judged recent trends in French government policy and spending to be positive. Even so, it says that a number of recent French initiatives have been too timid, and that others may have been hindered by an increased fiscal burden on industry to support the government's social policies.

OECD also criticizes French industry for its attitudes to research. The report quotes a survey which showed that 68 per cent of managers and engineers in French industry think that research is not an essential part of innovation: and that 78 per cent would seek potential innovators in the *grandes écoles*, the elite and dustily academic French engineering schools, rather than in the universities (where most French research is done).

The "ivy league" of French higher education and its graduates are therefore out of touch with the needs of modern innovators, OECD judges, but while recognizing the problem, the present socialist government has been meek in its attempts to solve it.

The OECD authors lay a great deal of blame on the education system, and on the schools which are largely geared to nurturing the narrow cleverness that the *grandes écoles* require from their entrants. French teaching remains "purely deductive and abstract" and this "has prevented France, unlike Japan, the United States and Germany, from developing an authentic industrial and enterprise-based culture", the reports claim.

Nevertheless the present government has taken some action in these areas, and the report praises efforts to extend the

granting of a *baccalauréat* (high school diploma) and hence status to a wider range of technical subjects. Efforts to attract more of the *grandes écoles* graduates, who generally end up in high management or civil service positions, into taking a research degree are also praised, though the numbers involved are small (500 a year at present).

OECD also gives a pat on the back to the ministry of education for attempting to give the universities more serious structures of evaluation and a concomitant autonomy. But "it seems so far as higher education is concerned things have come to a stop half way along the road, hesitating to introduce any real autonomy".

Research councils, such as the Centre National de la Recherche Scientifique (CNRS), with its 25,000 staff and 1,300 laboratories (two-thirds of which are run jointly with the universities), have increased their power, according to OECD but have simultaneously made efforts at greater consultation and regionalization of policy-making. But OECD criticizes the influence it believes is still exerted by a whole series of pressure groups that are "profoundly conservative — that is, most of the unions of teachers and researchers and other bodies". The research councils, despite their apparent formal power, are "relatively impotent in the face of an egalitarian grass-roots attitude opposed to clear-cut decisions", is the report's view.

The reforms in research policy under Mitterrand have, however, been "efficient and tailored to the French system". Their greatest success has been to create "a consensus round national science policy", and to encourage an outward-looking shift in the attitudes of researchers towards industry and other social partners such as the schools and regional government.

But altogether the French system is still riven by a system of contradictory "egalitarian and feudal" tendencies that hinder change, say the authors of the OECD report. Before the climate for innovation in France can really improve, major long-term educational and social changes will have to take place, or be engineered, the report concludes. But there is much of France that will have heard all this before. Whether the changes are desirable, and whether they will be more likely to take place under the present, or a new, government is for the French electorate to decide.

Robert Walgate

Reviews of innovation policies: France by Viscount Etienne Davignon (former vice-chairman of the European Commission), Professor Umberto Colombo (president of the Italian Nuclear and Alternative Energy Agency, ENEA), Dr A.P. Speiser (director of research, Brown Boveri and Co. Ltd and chairman of the research and development committee of the Swiss Confederation of Industry), and Professor J. Zysman (Institute of International Studies, University of California, Berkeley). OECD Directorate for Science, Technology and Industry (Paris 1986).

Plant gene conservation

SIR—In recent months, controversy has arisen internationally over the conservation and availability of plant genetic resources, the genetic diversity of primitive varieties of crop plants and related wild species upon which future crop improvement programmes will depend.

Much of the controversy has centred upon the alleged exploitation by "gene-poor" countries of the "North" of the "gene-rich" countries of the "South", and upon the role of the International Board of Plant Genetic Resources (IBPGR) in genetic conservation programmes. The loss of genetic diversity contained in traditional varieties through their replacement by modern ones has occurred in most parts of the world. Many scientists worldwide have attempted to preserve this diversity by collecting and conserving it. They have been greatly supported by the efforts of IBPGR, which has catalysed activities and provided funds for their execution.

It is clear that, without adequate conservation strategies, many genetic resources would quickly be lost. One of the most effective ways is to store seed in gene banks, in which seeds are dried to a low moisture content and stored at low temperatures. Not all seeds can be stored in this way and, as with vegetatively propagated plant species, alternative methods such as tissue culture (*in vitro* culture) have been devised.

Three aspects of genetic resources work are now being called into question by various people, including the Canadian agricultural economist Pat Mooney. The first is that the countries of the north have "robbed" the southern countries of their genetic resources patrimony and are reluctant to return it to them. The second is that it is unwise to store genetic resources in seed banks because selection encourages the survival of genotypes best suited to survive under such conditions. The third is that IBPGR is supporting developed countries at the expense of the developing ones.

None of these allegations can be substantiated. Firstly, are the developed countries restricting the free flow of germplasm to the developing ones? We can find very little evidence of this. By and large, the record of northern countries in providing material for the southern ones is excellent. Thus the US Department of Agriculture (USDA) small grains seed bank at Beltsville, Maryland, has virtually never denied material to any country; not only basic germplasm but commercial varieties have always been made freely available. In 1983 alone, 150,000 samples of wheat and wheat relatives were distributed, and more than 50 per cent of these consignments were to overseas countries, mostly in the developing world. The same is true for countries in Europe that have gene banks: when available, seed is always sent to every *bona fide* scientist requesting it.

It is true that there have in the past been more seed banks in the developed countries than in the developing countries. This is because the whole concept of genetic conservation was formulated initially in the Soviet Union with Vavilov's work, and afterwards in the United Kingdom, Australia and the United States, with the Food and Agriculture Organisation of the United Nations (FAO) playing an important role.

The seed banks were also first established in the industrial countries because they possessed the technical knowhow and scientific resources.

Nevertheless, the industrial nations have financed the biggest operation the world has ever known to preserve genetic resources, and this they have done through special funding to IBPGR and the International Agricultural Research Centres (IARCs).

The second criticism questions whether genetic resources ought to be conserved in seed banks. There is now a considerable body of scientific data which clearly demonstrates that under good storage conditions seeds can be stored for 30–50 years and possibly very much longer with little loss of viability or genetic erosion. In the absence of good storage conditions, the viability falls quickly and the germplasm is lost. Even under ideal storage conditions, the seeds must eventually be "regenerated" by growing out a new set of plants and obtaining fresh seed from them. If this is done carefully, selection can be kept to a minimum and the samples will not, as is alleged, become adapted to survive only in seed bank conditions.

Even allowing that a certain low level of selection may take place in seed banks, what is the alternative? Wild crop relatives may be conserved in nature reserves in some cases but, at present, there are not enough of these to deal with more than a minute fraction of the problem. Seeds of old crop varieties, it has been suggested, might be kept in special farm reserves in their places of origin. This is quite impracticable. There are some 130,000 accessions of wheat, 83,000 of rice, 14,000 of barley and at least 30,000 of maize in various gene banks, probably more. It would be quite impossible to grow these in their original areas, given the pressing need for land and food among developing countries. This land is needed for growing higher yielding, better adapted and more resistant varieties, which themselves are largely the result of the international use of freely available germplasm. There is thus an overwhelming case for the continuation of seed banks working to standards established by international committees set up by IBPGR. Such standards have been accepted by scientists all over the world. IBPGR has also recently established the basic procedures for *in vitro* gene banks, again working with distinguished scientists from northern and southern countries.

Finally we should like to counter the criticism aimed at IBPGR, that it is largely concerned with helping the developed countries. Nothing could be further from the truth.

IBPGR was established in 1974 as an autonomous international scientific organization under the auspices of the Consultative Group on International Agricultural Research (CGIAR). This latter body mobilizes financial support from donor countries and foundations to meet the budgetary requirements not only of IBPGR but also of the network of IARCs such as the International Rice Research Institute (IRRI) in the Philippines, and others dealing with a wide range of crops. All these centres have germplasm programmes, storing genetic resources of their respective crops mainly as seeds but also, where appropriate, *in vitro* conservation is being developed. Their germplasm work has received considerable financial support from IBPGR, and they have used this germplasm to provide improved breeding material for national breeding programmes in developing countries.

IBPGR also supports national genetic resources efforts, has financed and encouraged work in the developing world, and has trained numerous developing country scientists. It has financed 300 collecting missions in approximately 90 countries and provided basic equipment for the national gene banks in 28 developing countries. In respect of storage facilities, IBPGR has reached agreement with 31 countries for long-term storage, of which 25 are located in developing countries. As 80 per cent of all these long-term germplasm storage centres are in developing countries, those who assert that IBPGR intends that the developed countries should continue to hold most of the world's germplasm are clearly mistaken. What developed countries have accomplished is to collect and hold in trust the genetic resources of the developing countries. They have now returned a considerable part of these and are only too happy to return any or all of the rest that the developing countries may desire.

IBPGR, in common with the CGIAR system of IARCs, is promoting active use of this germplasm, through schemes for its evaluation and development into better adapted, higher yielding, more nutritious and more disease-resistant varieties for developing countries.

We conclude that genetic resources conservation is in competent hands and should remain there. What is becoming abundantly clear is that the germplasm policies of IBPGR, the IARCs and CGIAR are producing results through wise allocation of funding and genetic resources materials. Where the infrastructure is present, these policies pay off handsomely, thus providing the potential for improved well-being and economic advancement in the developing countries.

M.H. ARNOLD (Plant Breeding Institute, Cambridge), D. ASTLEY (National Vegetable Research Station, Warwick), E.A. BELL (Royal Botanic Gardens, Kew), J.K.A. BLEASDALE (National Vegetable Research Station, Warwick), A.H. BUNTING (University of Reading), J. BURLEY (University of Oxford), J.A. CALLOW (University of Birmingham), J.P. COOPER (Stroud, Gloucestershire), P.R. DAY (Plant Breeding Institute, Cambridge), R.H. ELLIS (University of Reading), B.V. FORD-LLOYD (University of Birmingham), R.J. GILES (Plant Breeding Institute, Cambridge), J.G. HAWKES (University of Birmingham), J.D. HAYES (University College of Wales, Aberystwyth), G.G. HENSHAW (University of Bath), J. HESLOP-HARRISON (Welsh Plant Breeding Station, Aberystwyth), V.H. HEYWOOD (University of Reading), N.L. INNES (Scottish Crop Research Institute, Dundee), M.T. JACKSON (University of Birmingham), G. JENKINS (Plant Breeding Institute, Cambridge), M.J. LAWRENCE (University of Birmingham), R.N. LESTER (University of Birmingham), P. MATTHEWS (John Innes Institute, Norwich), P.M. MUMFORD (University of Birmingham), E.H. ROBERTS (University of Reading), N.W. SIMMONDS (University of Edinburgh), J. SMARTT (University of Southampton), R.D. SMITH (Royal Botanic Gardens, Kew), B. TYLER (Welsh Plant Breeding Station, Aberystwyth), R. WATKINS (East Malling Research Station, Maidstone), T.C. WHITMORE (University of Oxford), L.A. WITHERS (University of Nottingham)

Darwin, Freud and creationism

SIR—I would like to comment on Anthony Clare's approbatory review (*Nature* 318, 112; 1985) of a number of books purporting to dispose once and for all of Freud's pretensions as a scientist, by drawing a comparison with the creationists who would have us abandon Darwin's contributions to biology. Neither Darwin nor Freud would have claimed in their time that their subjects had reached what Kant called the sure path of a science (though it could be thought that molecular biology has more recently provided the intellectual key to an adequate theory of evolution); but to dismiss Freud's attempt to grasp the nature of the problem presented to us by human behaviour and its aberrations because, like Darwin, he needed to substitute metaphor for mechanism in his preliminary sketch of what might become a rational psychology, is to throw out the baby in order to play with the bath water. If psychoanalysis is to be regarded as of no account, what kind of psychiatry are we to be left with? And if what we are left with is all we are going to get, how are we to respect its protagonists as critical scientists when they are content with so unsatisfactory an amalgam of empiric treatments and shaky statistics (the latter useful as a screening test to disprove some tentative hypotheses but unable, for reasons that Karl Popper has given us, to generate alternative ones) as the basis for their surely not very satisfactory practice? As Sir George Pickering (no friend to analysis) once said to me, an ounce of honest if mistaken endeavour to understand a problem is worth a ton of destructive criticism; for truth remains the child of error, not confusion.

JOHN A. DAVIS

University of Cambridge Clinical School,
Department of Paediatrics,
Level 8, Addenbrooke's Hospital,
Hills Road, Cambridge CB2 2QQ, UK

PhD theses

SIR—P. A. Lund (*Nature* 317, 470; 1985) writes about the "labour pains involved in the production . . . [of a] PhD thesis". The major problems in most of the graduate schools that I have been associated with in Europe, Canada and the United States are: (1) the production of trivial PhDs; (2) the "bogging down" of many PhD students by insurmountable experimental or theoretical difficulties; (3) the virtual abandonment of some PhD students, or in many cases lack of communication, between them and their supervisors.

At the Feinberg Graduate School of this institute (where we have some 500 MSc and PhD students), we initiated a system

many years ago (see D. Samuel *Pure and Applied Chem.* 22, 163; 1970) for raising the standard of PhDs and improving the quality of supervision. Within a year of being admitted, each PhD student must submit a detailed written research proposal, based on his reading of the published literature on the subject and preliminary laboratory work or calculations — in effect, a "feasibility study". This proposal is then sent to two senior members of the staff, not necessarily precisely in the same field, who interview the student and determine his overall knowledge, and his ability to undertake the proposed research project and complete it in a reasonable time.

The staff members continue to see the student and periodic reports, throughout his or her time at the institute. In spite of some initial opposition to this scheme, on the grounds of restricting the "academic freedom" of the supervisor, as well as the time involved in following the career of each student, it has worked extremely well for over twenty years. Apart from raising the standard of PhDs and solving both scientific and personal problems, it has frequently helped in the research itself by encouraging an exchange of ideas between different research groups and departments. This scheme, and variations of it, are used in other institutes and universities but not widely in the United Kingdom. I would suggest that it be adopted more extensively in order to protect and encourage that very precious commodity, the graduate student, as well as producing, perhaps, better research.

SAMUEL

(Viscount Samuel)

Weizmann Institute of Science,
Rehovot,
Israel

SIR—Recent correspondence in *Nature* on the value of PhD theses suggests to me that many scientists do not realize that such theses are in fact readily available through university libraries. One important source is the thesis filming operation at the British Library Lending Division at Leamington Spa. In 1984–85, the British Library Lending Division's total expenditure for literature was more than £3,600,000 of which £86,000 was spent on dissertations. With the exception of London and the Council for National Academic Awards, all British universities submit their doctoral theses for copying. The British Library Lending Division holds 1,410 miles of roll microfilm. In 1984–85, 5,500 theses were filmed. Theses are sold through the British Library Lending Division by filling in a Thesis Declaration Form. Theses can be supplied on microfilm, photocopy or microfiche. On the whole, masters theses are not collected. The division buys, on demand, doctoral dissertations listed in *Dissertation Abstracts International* which are part

of the University Microfilms doctoral programme. Various indexes are available for identifying theses, including *Dissertation Abstracts International*, *British Reports, Translations and Theses*, and *Aslib's Annual Index to Theses Accepted for Higher Degrees by the Universities of Great Britain and Northern Ireland and the Council for National Academic Awards*.

SHEILA M. MOULD

(Librarian)

MRC Clinical and Population

Cytogenetics Unit,
Western General Hospital,
Edinburgh, EH4 2XU, UK

Neptune's rings

SIR—Your leading article "Whose rings around Neptune?" (*Nature* 318, 505; 1985) contains some errors. First, proper credit for the observations made at Cerro Tololo would mention that I was the visiting astronomer who made the observation of the occultation by the ring arc, with the assistance of L.R. Elicer (whose help greatly streamlined the observing procedure). Second, the observations were not made "in some haste" as stated. We observed both the star and planet on the night of 21 July with different filters in order to select three filters that would give good signal-to-noise across a broad spectral range using the 0.9-m three channel photometer. I chose to use the Johnson U (0.44 μm), Johnson V (0.55 μm) and an infrared (0.87 μm) filters. Both the Johnson U and IR filters were internally mounted in the photometer. The Johnson V filter was mounted in a movable filter slide with a handle extending from the photometer very near to where the astronomer is positioned to guide the telescope. We suspect that one of us knocked the slide so that the occultation was observed through another filter, linear combination of filters or combination of filter and open sky, as the counts did not correspond to the test of the visual wavelength filter counts we conducted at the beginning of the night on 22 July.

Finally, Lissauer's article could have referred to an abstract in the Lunar and Planetary Conference XVI, p. 368, 1985, which includes all of the scientists involved and all of the data combined together. (Although I understand that Manfroid and Haefner requested that their names and data be withdrawn from the abstract, the volume of abstracts had already been printed when this request was made.)

I would like to add that William Hubbard and I have continued a productive and cooperative affiliation with the French astronomers for subsequent observations of occultations by the Neptune system.

FAITH VILAS

NASA, Lyndon B. Johnson Space Center,
Houston, Texas 77058, USA

Frost resistance and *Pseudomonas*

A professor of biophysics and medical physics at the University of California, Berkeley, reflects on the arguments that have halted the ice-minus tests.

FOR many decades, plant geneticists have worked to develop frost-resistant strains and varieties of crop plants. In this general endeavour, a novel weapon has recently been proposed: the dispersal of strains of bacteria from which the gene for ice nucleation protein¹ has been removed. The bacteria are *Pseudomonas syringae* and *Pseudomonas fluorescens*, which occur naturally in forms that are ice-nucleation-active (INA), containing the functional gene, and ice-minus (INA⁻).

Pseudomonads are non-spore-forming psychrophilic saprophytes, universally dispersed in soil, water and food. Some cause plant diseases. One widespread species, *Pseudomonas aeruginosa*, is occasionally infective in animals, but such infections are rare. *Ps. syringae* does not grow at 37°C and is non-infective. Other species are familiar in food spoilage, causing, for example, the earthy and ammoniacal flavour of stale meat.

A field test of INA⁻ *Ps. syringae* and *fluorescens*, applied to 2,400 blossoming strawberry plants, has been scheduled by Advanced Genetic Sciences (AGS), a biotechnology company². The test would, it is hoped, pre-empt the ice-formers by the ice-minus strain. Approval of this field experiment was given by the US Environmental Protection Agency after prolonged deliberation and the inclusion of extensive safeguards.

The length of the growing period, which decides whether a crop can be produced, is the number of days between the last and first killing frosts in spring and autumn respectively. Data are compiled by the US Department of Agriculture for every county in the United States³. Frost prevents dent corn (maize) from ripening in northern latitudes, notably in Canada.

The search for frost resistance is not new, but was dramatized by the spurious claims of the Soviet charlatan Lysenko, who asserted in the 1950s that his methods would make it possible to produce frost-resistant wheat. This led in Siberia "to extremely low yields and sometimes complete destruction of the crop by frost"⁴. But the greatest dramatization of frost resistance was a reality. Canadian plant breeders produced a new strain of wheat, named Marquis, in 1904. Three years later, in Saskatchewan, Marquis wheat plants in a small experimental plot survived a killing frost on 12 September 1907⁵. As a result, Marquis pushed wheat farming northward into the rich black soil

of the prairie provinces, previously untiled, and Canada was able to send millions of bushels of wheat to China in the 1950s.

I looked forward to news of ice-minus *Pseudomonas*. To my surprise, I learned that a protest demonstration would take place on 15 January on my doorstep. (I work on the fourth floor of a University of California building, of which the third floor is leased to AGS.) The demonstrators, organized by "East Bay Greens" and "Earth First!", distributed a leaflet couched in inflammatory terms alleging that "this genetically altered bacteria (*sic*), if established on Kudzu, could render thousands of acres of agricultural land virtually useless within a few years, and eventually dominate the eastern landscape". I pointed out to a demonstrator that, if the Kudzu plant, an "escaped" weed of Asian origin, could be so easily frost-proofed, the same could happen to tomato plants in Montana and orange trees in Florida, but I was informed that too much food was being produced anyway. The leaflet also said that once they escape, one-celled organisms can cause more death and destruction than all the wars we have ever fought, and it continued:

We oppose the release of these genetically altered organisms into the environment. We oppose the research that has created this technology as morally bankrupt and motivated by higher profits... It is not for our scientists, bureaucrats and industrialists to play God, or endanger us with technology that no one understands. While this experiment may prove safe in the end, what of the many thousands that are sure to follow, and the few that ultimately fail? The risks are not being adequately addressed, nor are the pressing moral issues that this type of technology raises.

I commented to an interlocutor that one-celled organisms always escape, that the profit motive in agriculture often leads to development of new technology that helps feed people everywhere, that the organisms did not effectively or essentially differ from ice-minus strains occurring normally in the environment, that this experiment did not involve "technology that no one understands" and that each experiment should be judged on its own merits rather than on the basis of "many thousands that are sure to follow". I also remarked that aiding the production of food was a moral issue. But the demonstrators then started to chant slogans and wave placards for the benefit of the television cameras and news reporters. One

placard stated, "We don't need no designer genes". A leader of the group read a telegram from 27 German Bundestag members alleging that "our health and environment must not be sacrificed".

I was given a leaflet saying that "Green politics" seeks a "humane, just, democratic, peaceful, and ecologically sustainable world for ourselves and our children", with which I agree. But I do not understand why what seems to be harmless procedure for increasing the yields of produce is incompatible with this goal, especially since we are told to eat more fruit and vegetables to protect the health of ourselves and our children. But apparently there is no relenting.

In a local newspaper report on the demonstration, a representative of "Earth First!" stated that "This is very similar to the first atmospheric test of the atom bomb. Nuclear physicists who created the bomb had the feeling that their actions could possibly ring down the curtain on humanity"⁶. Since a bucketful of ordinary dirt contains as many pseudomonads as would be used in the proposed experiment, either our species will disappear in the next minor dust storm, or the statement is nonsensical, probably the latter.

The hearings in Salinas, Monterey County, on 27 January, predictably led to a "postponement" of the scheduled test. According to *USA Today*, "school trustee Judy Pennycook said: 'There is no way north Monterey County can be considered remote. There are 30,000 people there', and 'We're playing environmental roulette', said tree farmer Glenn Church." According to Jeremy Rifkin, who is busily filing suits to block the experiments, the modified bacteria "may decrease rainfall". But it would seem that naturally occurring INA⁻ are currently in equilibrium with the INA forms, and Rifkin's assertion was contradicted by R. Schnell, a meteorologist.

I perceive a ritualistic rather than scientific attitude by opponents of the test, and I see no tendency by them to consider benefits as weighed against what seems to be a well-evaluated and minuscule risk.

Thomas H. Jukes

1. Green, R.L. & Warren, G.J. *Nature* **317**, 645-648 (1985).
2. Palca, J. *Nature* **319**, 284 (1986).
3. *Climate and Man*, USDA Yearbook of Agriculture (Superintendent of Documents, Washington, DC, 1941).
4. Medvedev, Z.A. *The Rise and Fall of T.D. Lysenko* (Columbia University Press, New York, 1969).
5. deKruif, P. *Hunger Fighters* (Harcourt, Brace, Jovanovich, New York, 1928).
6. *Grass Roots* **14**, 1, 29 January - 11 February (2002 Blake Street, Berkeley, California 94704, 1986).

Origin of life

The RNA world

from Walter Gilbert

UNTIL recently, when one thought of the varied molecular processes at the origin of life, one imagined that the first self-replicating systems consisted of both RNA and protein. RNA served to hold information, whereas protein molecules provided all the enzymic activities needed to make copies of RNA and to reproduce themselves. The cycle that developed a self-replicating system out of the primitive soup of amino acids and nucleotides had two radically different components¹.

Now it seems possible that the informational and catalytic properties of these two components may be combined in a single molecular species. Last week in these columns Frank Westheimer² described the discovery of enzymic activities in the RNA molecules of *Escherichia coli*, in which ribonuclease-P cuts phosphodiester bonds during the maturation of the transfer RNA molecule^{3,4}, and of *Tetrahymena*, whose ribosomal RNA contains a self-splicing exon^{5,7}. If there are two enzymic activities associated with RNA, there may be more. And if there are activities among these RNA enzymes, or ribozymes, that can catalyse the synthesis of a new RNA molecule from precursors and an RNA template, then there is no need for protein enzymes at the beginning of evolution. One can contemplate an RNA world, containing only RNA molecules that serve to catalyse the synthesis of themselves⁶.

The self-splicing intron is an RNA element that can splice itself out of an RNA molecule. This reaction should be reversible, and the intron could splice itself back into an appropriate nucleotide sequence. Thus, in the RNA world, such introns could both remove and insert themselves into the background of replicating RNA molecules. The significance of this is not the simple insertion and removal of introns, but the fact that two introns, separated from each other by another RNA element, an exon, can combine with each other so as to remove as a unit both themselves and the intervening exon from one RNA molecule and to insert into another. Thus, self-inserting introns can create transposons to move exons around. This property provides RNA with a major evolutionary facility that it otherwise lacks — recombination, the ability to produce new combinations of genes. Of course the self-replicating molecules would in any case have evolved slowly by miscopying, that is, by mutation. But transposons provide the equivalent of sex — the infectious transmission of genetic elements from one organism to another. Recombination and sex are powerful devices to permit a more

useful exon to pass from one replicating structure to an unrelated one.

This picture of the RNA world is one of replicating molecules that reassort exons by transposable elements created by introns. This process builds and remakes RNA molecules by chunks and also permits the useful distinction between information and function. Information storage needs to be one-dimensional, for ease of copying, but molecules with enzymic functions tend to be tight three-dimensional structures, whose forms are unrelated to the demands of any copying mechanism. (This dichotomy is most obvious today between the linear order along DNA and the structure of proteins.) In the RNA world, the structure that would be replicated has the full complement of introns. Some of the daughters, by splicing out all their introns, would convert to functional molecules, the ribozymes. A remnant of this process may be the structure of transfer RNA, where a compact secondary structure is broken up by the insertion of an intron.

The first stage of evolution proceeds, then, by RNA molecules performing the catalytic activities necessary to assemble themselves from a nucleotide soup. The RNA molecules evolve in self-replicating patterns, using recombination and mutation to explore new functions and to adapt to new niches. By using RNA cofactors, such as nicotinamide adenine dinucleotide and flavin mononucleotide they then develop an entire range of enzymic activities⁹. At the next stage, RNA molecules began to synthesize proteins, first by developing RNA adapter molecules that can bind activated amino acids and then

by arranging them according to an RNA template using other RNA molecules such as the RNA core of the ribosome. This process would make the first proteins, which would simply be better enzymes than their RNA counterparts. I suggest that protein molecules do not carry out enzymic reactions of a different nature from RNA molecules but are able to perform the same reactions more effectively and rapidly, and hence will eventually dominate. These protein enzymes are encoded by RNA exons, thus they, in turn, are built up of mini-elements of structure.

Finally, DNA appeared on the scene, the ultimate holder of information copied from the genetic RNA molecules by reverse transcription. After double-stranded DNA evolved there exists a stable linear information store, error-correcting because of its double-stranded structure but still capable of mutation and recombination. RNA is then relegated to the intermediate role that it has today — no longer the centre of the stage, displaced by DNA and the more effective protein enzymes. But a few RNA enzymic activities still exist, the two described recently³⁻⁷, and possibly others in the role of ribosomal RNA or in the splicing of eukaryotic messenger RNA. The relic of this process is the intron/exon structure of genes, left imprinted on DNA from the RNA molecules that earlier encoded proteins, a residue of the basic mechanism of RNA recombination. □

1. Eigen, M., Gardiner, W., Schuster, P. & Winkler-Oswatitsch, R. *Sci. Am.* **244**, 88 (1981).
2. Westheimer, F.H. *Nature* **319**, 534 (1986).
3. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. *Cell* **35**, 849 (1983).
4. Guerrier-Takada, C., & Altman, S. *Science* **223**, 285 (1984).
5. Kruger, K. *et al.* *Cell* **31**, 147 (1982).
6. Cech, T.R. *Int. Rev. Cytol.* **93**, 3 (1985).
7. Zaug, A.J. & Cech, T.R. *Science* **231**, 470 (1986).
8. Sharp, P.A. *Cell* **42**, 397 (1985).
9. White, H.B. III *J. molec. Evol.* **7**, 101 (1976).

Walter Gilbert is at the Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

Infrared detectors

Superlattices point ahead

from Gordon C. Osbourn

TECHNICAL developments in the past few years have made clear what the next generation of infrared detectors will be like. The need is for semiconductor materials appropriate for long-wavelength ($\geq 12 \mu\text{m}$) imaging, at the focal planes of new telescopes and for various defence applications. Most effort has focused on the II-VI alloy system (Hg,Cd)Te, but work has also begun on a recently proposed III-V system consisting of many alternating layers of thin mismatched In(As,Sb) crystal layers, called strained-layer superlattices (SLSs). A crucial first step in their

development has now been achieved with the successful growth of crystals of SLSs in the In(As,Sb) alloy system^{1,2}.

The materials from which long-wavelength detectors might be made are limited by the stringency of the requirements. These imaging devices will consist of two-dimensional arrays of many individual photovoltaic detector elements on a single wafer of a material which must have an energy bandgap less than or equal to 0.1 eV at about 77 K, the operating temperature of the infrared detector. This bandgap is smaller than that of all the III-V alloy

materials in homogenous form. Alloys of II–VI composition, such as (Hg,Cd)Te with approximately 80 per cent HgTe, do have the required low bandgaps, which is why they have been studied intensively. But another requirement for the infrared material is that the bandgap value should be very uniform across the wafer, so that most of the individual photovoltaic detector elements can function properly. For this reason it will be necessary to control (Hg,Cd)Te compositional variations to much less than 1 per cent over large areas. This requirement is very difficult to meet with existing growth techniques, causing concern about the yield of devices based on (Hg,Cd)Te alloys and their cost. Materials for infrared detection must also be stable and able to survive severe conditions at some processing steps and harsh environments in operation. Their bond strengths are generally low because of the large atomic numbers of the constituent atoms. (Hg,Cd)Te is particularly brittle on this account and because of the significant ionic component of the bonding.

Although the bandgaps of homogenous III–V alloys are larger, In(As,Sb) SLSs are expected to have the desired 0.1 eV value¹. This comes about because of the alterations of layers in compression and tension; the larger layer strains needed to accommodate the differences in atomic spacing of the SLS layers act to reduce the bulk bandgap of those layers in the III–V SLS which are in tension. SLSs with layer strains greater than or equal to 0.5 per cent can show strain shifts which are enough to bring the gap of the In(As,Sb) alloy with the lowest gap value (60 per cent InSb) down to 0.1 eV at 77 K.

In(As,Sb) SLSs are expected to have some valuable features for detector arrays. For example, their bandgaps are fairly insensitive to compositional variations in the per cent range. Such variations in the In(As,Sb) layer compositions across the SLS wafer should be tolerable for detector applications. Moreover, it is expected that the III–V In(As,Sb) SLSs will have greater bond strengths than the (Hg,Cd)Te alloys suitable at long wavelength because of the larger covalent contribution to their bonding.

Some of these predictions have now been put to the test following the first successful growth of In(As,Sb)-based SLSs^{2,4}. X-ray diffraction and transmission electron microscopy studies have verified the presence of the alternating SLS layers, and bond strengths in these materials seem sufficient to allow the growth of the thin superlattice layers at temperatures of 400–500°C for several hours without any large-scale interdiffusion of the layers.

Typical overall structures consist of InSb substrates (a few hundred microns thick); an InAs_{1-x}Sb_x 'buffer layer' which is first grown on the substrate; and finally a

superlattice (~1 µm thick) which consists of many thin alternating strained layers of InAs_xSb_{1-x} and InAs_ySb_{1-y}, where x and y are different.

Defects are a problem. The buffers typically contain many lattice defects. In previous SLS studies with larger bandgap materials, it has been found that the strained layers prevent buffer defects from penetrating through the superlattice, but initial electron microscope studies surprisingly show this not to be the case for the In(As,Sb) SLSs. Indeed, many buffer defects are seen to propagate through the SLS, and will hamper detector function unless they can be eliminated.

The reasons for the prevalence and persistence of these defects are not yet established, but the main focus of current work on In(As,Sb) SLSs is to eliminate buffer defect propagation. Some approaches include the use of graded composition buffers; buffers with less mismatch relative to InSb; or alternative substrates with buffers in compression.

Meanwhile, attention continues to be focused on the (Hg,Cd)Te alloys. One promising approach is to use superlattices made of alternating layers of the II–VI binaries HgTe and CdTe^{5,7}, which have been shown to have the desired bandgap value. One advantage of this system is that bandgap variations across the superlattice wafer arise only from layer-thickness variations which can be very precisely controlled in growth by molecular beam epitaxy. The most serious difficulty in the approach is that there may be significant layer interdiffusion even at the very low growth temperatures (185°C) used⁸. Thus, superlattices in the (Hg,Cd)Te system are likely to be similar to the (Hg,Cd)Te alloys themselves with respect to stability during device processing and in harsh environments.

The recent trend in infrared materials research has been in the direction of artificially layered structures and new layer materials. The new-found flexibility in designing and controlling the properties of novel superlattice materials promise to help in meeting the difficult requirements for infrared focal plane arrays. But the physical properties of these layered materials are interesting in themselves and so may well provide further possibilities. □

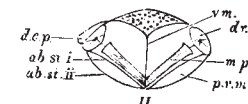
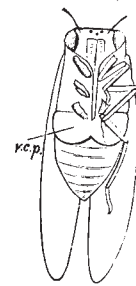
1. Lee, G.S. *et al.* *Appl. Phys. Lett.* **47**, 1219 (1985).
2. Dawson, L.R. *Molecular Beam Workshop*, Minneapolis (September 1985).
3. Biefeld, R.M. *Conf. Cryst. Growth* (Am. Ass. Cryst. Growth, Stanford, 1985).
4. Osbourn, G.C. *J. Vac. Sci. Tech.* **B2**, 176 (1984).
5. Schulman, J.N. & McGill, T.C. *Appl. Phys. Lett.* **34**, 663 (1979).
6. Smith, D.L., McGill, T.C. & Schulman, J.N. *Appl. Phys. Lett.* **43**, 180 (1983).
7. Faurie, J.P., Million, A. & Piagnet, J. *Appl. Phys. Lett.* **41**, 713 (1982).
8. Arch, D.K. *et al.* *Phys. Chem. HgCdTe Workshop* (San Diego, 1985).

Gordon C. Osbourn is Supervisor of Electronics and Transport Phenomena in Solids, Sandia National Laboratories, Albuquerque, New Mexico 87185, USA.



100 years ago

The Singerjie (*Platyleura capensis*) is a well-known insect at the Cape; and few visitors to the shores of Table Bay can have failed to notice, in the hotter months of the year, the sharp shrill metallic sound produced by the "little singer." The male Cicada alone possesses the power of singing, the female — recognised at once by the long ovipositor folded beneath the abdominal somites — being dumb. If the ventral surface of a male singerjie be examined (Fig. I.) two large ventral cover plates (*v.c.p.*) are seen meeting in the central line and extending backwards from the metathorax over the anterior abdominal somite. On turning the insect over and looking at the dorsal surface, two very much smaller dorsal cover plates are seen extending forward from either side of the first abdominal somite. If one of these plates be removed there is seen the wrinkled surface of a thickish chitinous membrane, the drum. Turning the insect over again, and removing one of the ventral cover plates, two membranes are disclosed, separated by a transverse chitinous support. Of these the anterior is white, narrow, and opaque, the posterior (*p.r.m.*, Fig. II.) translucent, oval, and tightly stretched.



The transverse chitinous support (*ab.st.ii*) is the sternum of the first abdominal somite; the membranes would seem to be modified to act as resonators. The second ventral cover plate may now be removed, disclosing the anterior and posterior resonator membranes; the anterior membranes may be cut through; and the abdominal portion of the insect may be separated from the thorax. There are seen, taking their origin from the mid-line of the first abdominal sternum (Fig II., *ab.st.i*), two muscular pillars (*m.p.*), each of which, terminates in a chitinous plate, the upper surface of which is connected by a fine ligament with the drum (*dr.*). Under a low magnifying power this drum is strengthened with brownish ribs, which, together with its general elasticity, cause it to spring back after it has been drawn forward by the action of the muscular pillars, the fibres of which are beautifully striated. Each time the drum is drawn forward and springs back, by the alternate contraction and relaxation of the muscular pillars, a sharp click is heard, as may readily be proved experimentally on the dead insect. From *Nature* 33 369, 18 February 1886.

Erratum

In Frank Westheimer's article "Polyribonucleic acids as enzymes" (*Nature* 13 February, p.534), figures 1 and 2 were transposed although the legends were correct.

Immunology

Seeing the way to B-cell growth

from John C. Cambier

ANTIBODY produced by the differentiated progeny of B lymphocytes represents a main line of defence against potential pathogens (antigens) such as bacteria and viruses. For decades, immunologists have attempted to define the species which regulate growth and differentiation of B lymphocytes. These studies, which hold the potential to lead to new strategies for treatment of immune dysfunction, have led to the common perception that B-lymphocyte immune responses are regulated by a vast array of T-cell derived molecules — lymphokines — as well as antigen and macrophage products. Noma *et al.* report on page 640 of this issue¹ the molecular cloning of complementary DNA encoding a T-cell derived molecule that exhibits multiple B-cell-regulatory activities previously ascribed to three different lymphokines. In a paper soon to be published Lee *et al.* report similar findings². These reports will have considerable impact on our understanding of regulation of B-cell function by T-cell products.

The complexity of lymphokine regulation of B-cell function stands in stark contrast to that of T cells. T-cell growth and differentiation appear to be regulated primarily by antigen recognized in association with molecules of the major histocompatibility complex (MHC) and a single T-cell product, interleukin-2. Evidence

from studies of disparate B-lymphocyte responses, which are components of the ultimate immune response, suggests that multiple lymphokines regulate distinct aspects of B-lymphocyte growth and differentiation. Specifically, the definition of B-cell growth factor 1 (BCGF₁), was based on its ability to promote proliferation of B lymphoblasts stimulated by anti-immunoglobulin antibody³, while another interleukin, BCGF₂ (see box), promoted proliferation of B cells stimulated with dextran sulphate⁴. Lymphokines that promoted proliferating B lymphocytes to differentiate and secrete immunoglobulin M (ref. 5) and immunoglobulin G (refs 6,7), called BCDF_μ and BCDF_γ respectively, were also described. Finally, a lymphokine was found that stimulates all resting B lymphocytes to increase cell surface expression of class II MHC molecules but did not stimulate them to proliferate^{8,9}.

These findings and reports of additional modulatory effects of interleukin-2, interferon-γ and interleukin-1 on lymphocyte function, and others suggesting that supernatants of certain T-cell clones^{10,11} contain additional lymphokines that modulate B-cell function, have left the field in a very confused state. The situation is complicated further by the fact that the purity of these factors isolated from complex T-cell supernatants is in many cases

suspect and, with the exception of interleukin-2 and interferon-γ, cloned gene products exhibiting these functional activities did not, until recently, exist.

The situation was simplified somewhat when Ohara *et al.*¹² purified to homogeneity a T-cell derived protein of relative molecular mass (*M_r*) 20,000, which they termed BSF-1, that was active in several of the functional assays described previously. This protein exhibited BCGF₁ activity¹², BCDF_γ activity¹³ and the ability to induce expression of class II MHC molecules on resting B cells¹⁴. This suggested that the same or very closely related lymphokine(s) mediate the multiple activities.

The description in this issue¹ of the cloning of a gene that encodes a protein of *M_r* 20,000 which functions in these three assays is a landmark for several reasons. It demonstrates conclusively that B-cell modulatory activities previously ascribed to multiple species can be mediated by a single lymphokine. Most importantly, it suggests that a single lymphokine interacting with B cells in distinct differentiative states has different regulatory effects.

This, of course, has far-reaching implications in suggesting that a ligand interacting with its receptor on cells in different states results in activation (or inactivation) of different gene sets. This process could be accommodated by a receptor which switches second-messenger coupling during cell differentiation, or by some as yet undefined genetic mechanism perhaps involving alteration of accessibility of enhancer or promoter sequences during cell differentiation. An alternative explanation is that this lymphokine signals all B lymphocytes equivalently, making them competent to respond to other currently undefined regulatory species (perhaps autocrine) which direct the qualitatively different responses observed.

The availability of the cloned gene product described by Noma *et al.*¹ and Lee *et al.*² and isolated B-cell populations in defined differentiative states will allow dissection of this important question. □

Two distinct candidates for the title interleukin-4 (IL-4) have been launched on the largely unsuspecting public. One is the protein (or glycoprotein) whose DNA sequencing by Honjo's group and a group at DNAX is put into context by John Cambier in the accompanying article. The other candidate is a protein previously termed eosinophil differentiating factor but now shown by C.J. Sanderson *et al.* of the National Institute of Medical Research, London to be the same as the B-cell growth factor BCGF₂ (*Proc. natn. Acad. Sci. U.S.A.* 83, 437; 1986).

Both proteins have a 'right' to be called IL-4 in the sense that an interleukin is merely a substance that carries a message between leukocytes (white blood cells). And in both cases the argument for renaming the protein IL-4 is the need to replace the two or three different names under which the protein has been known, each name reflecting the biological activity through which the protein has been independently discovered and rediscovered. It was an equivalent argument that led to IL-3 being adopted as the

name for a protein involved in the regulation of haematopoiesis and which had been known by several different names reflecting its varied biological activities.

The pity is that IL-4 has now been suggested as the name for two completely different proteins. The clash was known to the authors in advance but an attempt to avoid it failed. One of the issues that came up was whether the title of interleukin should be bestowed on a substance before it has been fully characterized in terms of amino-acid sequence. Certainly there has been a tendency among endocrinologists to call a 'factor' a 'factor' until its full characterization, whereupon it can be given a specific hormonal name. But the short history of the other interleukins does not follow that pattern.

In the end, such matters of nomenclature are the business of committees. The question of which of the two candidates should be called IL-4 will be discussed and may well be decided during the International Congress of Immunology in July. The 'loser' will presumably acquire the name IL-5. Peter Newmark

1. Noma, Y. *et al.* *Nature* 319, 640 (1986).
2. Lee, F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
3. Howard, M. *et al.* *J. exp. Med.* 155, 914 (1982).
4. Swain, S.L. *et al.* *J. exp. Med.* 158, 822 (1983).
5. Brooks, K., Yuan, D., Uhr, J.W., Krammer, P.H. & Vitetta, E.S. *Nature* 302, 825 (1984).
6. Vitetta, E.S. *et al.* *Immun. Rev.* 78, 137 (1984).
7. Sideras, P., Bergstedt-Lindquist, S. & Seversen, E. *Eur. J. Immun.* 15, 593 (1986).
8. Roehm, N.W. *et al.* *J. exp. Med.* 160, 679 (1984).
9. Noelle, R., Krammer, P.H., Ohara, J., Uhr, J.W. & Vitetta, E.S. *Proc. natn. Acad. Sci. U.S.A.* 81, 6149 (1984).
10. Leclercq, L., Bismuth, G. & Theze, J. *Proc. natn. Acad. Sci. U.S.A.* 81, 6491 (1984).
11. Leclercq, L., Cambier, J.C., Mishaal, Z., Julius, M.H. & Theze, J. *J. Immun.* 136, 539 (1986).
12. Ohara, J., Lahet, I., Inman, J. & Paul, W.E. *J. Immun.* 135, 2518 (1985).
13. Vitetta, E.S. *et al.* *J. exp. Med.* 162, 1726 (1985).
14. Oliver, K., Noell, J., Uhr, J.W., Krammer, P.H. & Vitetta, E.S. *Proc. natn. Acad. Sci. U.S.A.* 82, 2465 (1985).

John C. Cambier is Associate Professor at the National Jewish Center for Immunology and Respiratory Medicine, and at the University of Colorado Health Sciences Center, Denver, Colorado 80206, USA.

Mathematics

Demystifying the monster

from Ian Stewart

THE monster is one of those things that crop up every so often in mathematics: a new gadget, something very exceptional and unique, which seems to hint at concealed depths, all seen through a glass darkly. It is the largest sporadic simple group, containing a staggering septemdecillion or so elements (see *Nature News and Views* 317, 12; 1985). Until recently its existence was only suspected, though the evidence was very compelling; it was eventually constructed by Robert Griess (*Inventiones Mathematicae* 69 1; 1982) in a *tour de force* of combinatorics, especially noteworthy for the absence of computer assistance. That the monster should be tractable at all was a surprise to many; nevertheless, anyone who studies Griess's work is likely to wish that a simpler description may eventually emerge. Just such a simplification has now been given by John Conway (*Inventiones Mathematicae* 79, 513; 1985), which also yields new insight into the monster's inner nature.

To follow the story it is necessary to understand what abstract algebraists do. The ingredients for abstract algebra are a set, together with some operations on that set which allow its elements to be combined in various ways. The resulting structure is described by some standard name, depending on the precise laws obeyed by those operations. Thus, a group is a set with a single operation, combining elements x and y to form a product xy , obeying several laws, the main one of which is the associative law $(xy)z = x(yz)$. For example, multiplication defines a group operation on the set of non-zero

real numbers. A vector space is a set equipped with two operations, addition and multiplication by a scalar, satisfying about eight laws. An example — in a sense the example — is euclidean space of n dimensions. An algebra is a vector space equipped with a third operation, multiplication, satisfying yet further laws. One example of an algebra is the set $\mathbb{C}$ of complex numbers, which is a two-dimensional vector space consisting of expressions $x+iy$, equipped with an associative multiplication defined by $i^2 = -1$. It is a commutative algebra, which means that one of its laws is that $xy = yx$. Another is the quaternions $\mathbb{H}$, a four-dimensional space formed from expressions $w+ix+jy+kz$ where $i^2 = j^2 = k^2 = -1$, $ij = k = -ji$, $jk = i = -kj$, $ki = j = -ik$. This is associative but not commutative. Non-associative algebras of various kinds abound, and the most useful types have well-developed theories of their own.

There are all sorts of ways of taking these structures and building others from them. Groups can contain subgroups, vector spaces subspaces, sub-algebras and algebras. Moreover, some constructions start with one type of structure and produce another. In particular, the set of all symmetries of a given algebra turns out to be a group. By a symmetry of an algebra is meant a transformation of the algebra (say x becomes x') which preserves the algebraic structure: $(xy)'$ is the same as $x'y'$, and $(x+y)'$ is the same as $x'+y'$. In other words, 'operate and transform' has the same effect as 'transform and operate'. Thus, $\mathbb{C}$ has only two symmetries. One is

the identity, $(x+y)' = x+iy$. The other is more interesting: it is complex conjugation $(x+iy)' = x-iy$. The quaternions have many more symmetries (in fact, infinitely many) giving rise to a much bigger group.

It is no surprise that the symmetries of an algebra form a group, because the *raison d'être* of the group concept is to formalize the notion of symmetry. The symmetries of anything form a group, which gives groups a very fundamental role in mathematics. Simple groups are the 'atoms' of symmetry: groups that cannot be split up by a kind of factorization process into smaller groups.

A major triumph of twentieth century mathematics has been the complete determination of all simple groups with a finite set of elements. These fall into many well-behaved families, together with exactly 26 curious exceptions called the sporadic groups, an unruly crew each marching to its own drummer. The sporadic groups are a real puzzle: we know what they are, but why are they? The monster has pride of place among the sporadics as the biggest, brashest and in many ways the most intriguing of a pretty weird bunch.

Griess constructed the monster as the group of symmetries of an algebra of 196,884 dimensions. This number is not an accident: it derives from the computation of the 'character table' of the monster by B. Fischer, D. Livingstone and M. Thorne (University of Birmingham, 1978). They proved that, if it exists at all, the smallest space on which the monster can act as transformations has 196,883 dimensions. Griess found it convenient to stick on one extra dimension to give the algebra a nicer structure.

Griess's algebra is very complicated and obeys neither the commutative nor the associative laws. Nevertheless, it has several pleasant properties (the most pleasant, naturally, being that its symmetry group is of monstrous proportions). One of the hardest parts of the proof is to show that this symmetry group is actually finite. (Recall that a much smaller algebra, the quaternions, has an infinite symmetry group.)

Conway's construction follows a similar strategy but uses different tactics, to some extent guided by Griess's pioneering work. It involves a whole horde of exceptional combinatorial objects: the Golay code from coding theory; a Steiner triple system $S(24,8,5)$ in which a set of 24 objects is divided up into 8-element subsets, such that each 5-element subset is contained in a unique 8-element one; the Leech lattice (an economical way to pack spheres in 24-dimensional space); and a curious structure of the loop type (a group lacking the associative property) devised by R. A. Parker. Somehow these exotic creatures, as well as several others, are all part of the same big picture, whatever that

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Rhombohedral crystals of $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}] \cdot 6\text{THF}$. A description of this remarkable new molecule appears on page 666 of this issue (Photograph, S.J. Lippard).

may be. This assortment of ingredients is combined in the mathematical equivalent of a witches' cauldron to produce a powerful potion, an algebra of 196,884 dimensions which is '99 per cent associative' (see below). Three groups of symmetries of this algebra are defined, related by a remarkable 'triviality' property of Parker's loop. The monster, like Gaul, is made up from these three parts. Its finiteness is almost a triviality. The proof is an intricate virtuoso performance, typical of its inventor, and it represents a major demystification of both the monster and the sporadics.

It also casts an interesting light on the traditional way to find 'interesting' algebras,

which is basically to state new systems of laws and to look for algebras that obey them. The algebra for the monster is not defined in this way — it is like a fine upstanding citizen given to minor criminal tendencies. A random triple (x, y, z) chosen from this algebra has about a 99 per cent chance of obeying the associative law, but the other 1 per cent flout it, apparently with a clear conscience. This curious and tantalizing fact suggests a re-examination of the principles by which interesting algebras are identified. Is the time ripe for a theory of 'almost laws' of algebra? □

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AI, UK.

Neurotransmitters

Modifying channel function

from Charles F. Stevens

THREE recent articles — one on page 670 of this issue¹ and two^{2,3} in an earlier issue — shed light on one of the brain's most remarkable features: its ability to modify itself. Neurotransmitters bound to the surface receptors on neurones can cause the electrical excitability of the cell to change. All three reports present evidence that a coupling protein, G protein, forms the link between surface receptors and the alteration of properties of ionic channels that produce a change in the excitability of the cell. Unlike the best studied neuromodulatory systems, the mechanism does not involve changes in the concentration of cyclic (c) AMP or other second messengers in the cytoplasm; this chemical signalling system is limited to messenger molecules diffusing within the surface membrane of the cell.

Computers are self-modifying in that they sometimes alter their own programs, but the self-modifiability of the brain goes much further: the hardware itself is modified. One neurone applies a neurotransmitter to the surface of another; this transmitter can cause some of the ion channels on the second neurone to be covalently modified. Because the computing properties of a neurone are determined by the characteristics of its various ion channels, one neurone can change the way a second processes information.

The 'usual' way to modify channels — that is, the one we know most about — is to phosphorylate them with cAMP-dependent protein kinase⁴. The cAMP cascade has five main steps: (1) neurotransmitter binds to the cell surface; (2) bound receptor activates the enzyme adenylate cyclase that in turn (3) makes cAMP from ATP; (4) this cAMP activates the kinase that (5) adds a phosphate group to specific sites on the channel. Once it is phosphorylated the channel

behaves differently; it may open less often or may stay open longer each time it is activated depending on what sort of channel it is.

G proteins form a family with at least four members: two of these, G_i and G_o (sometimes called N_i and N_o), either stimulate (G_i) or inhibit (G_o) the adenylate cyclase. They in turn are activated by neurotransmitters binding to other receptors. The G proteins, then, are the messengers that tell the cyclase how active it should be according to which neurotransmitter receptors are occupied.

Why are the neuromodulatory cascades so complicated? Three reasons, at least. First, the same basic cascade can be used in different cells with only small changes — for example, the specificity of the receptor can be changed while the remainder of the system is kept the same. Second, each step can amplify the signal: one receptor can activate many enzymes by using the G-protein intermediate. Finally, the multiple steps provide many places where regulatory systems can interact. For example, stimulation of an enzyme by one neurotransmitter through G_i can be antagonized by the action of a second neurotransmitter through G_o .

The three recent papers all show that neurotransmitters act through a G protein, but that the traditional cAMP system

is not involved. The two earlier studies^{2,3} demonstrated that acetylcholine, which causes inhibition of the heart, works through a G protein that turns on a special set of potassium channels. The article in this issue¹ shows that calcium channels in dorsal root ganglion neurones are turned off by noradrenaline and by γ -aminobutyric acid (usually an inhibitory neurotransmitter) acting through a G protein.

How are the channels turned on or off by G proteins? One speculative possibility is that the G protein works directly on the channel just as, in other instances, it works on enzymes such as adenylate cyclase. If a G protein can modify enzyme function, why not channel function? A second possibility, for which there is some preliminary evidence⁵, is that another regulatory cascade is operating, one that involves the protein C kinase, to phosphorylate channels. The protein C kinase cascade also uses one or more members of the G-protein family, but its signalling molecule would be, in this context, diacylglycerol produced by a phospholipase, rather than cAMP made by adenylate cyclase as in the cAMP cascade. Whether the mechanism involves the direct action of G proteins on the channels or phosphorylation of channels by the protein C kinase (or some yet more mysterious alternative) must be decided by additional experiments.

When it was believed that neuromodulation generally worked by phosphorylation through the cAMP cascade everything seemed relatively simple, but now the story is getting more complex. Probably this is just the start: phosphorylation by many different kinases, direct regulation by G proteins, different second-messenger systems and regulation through GTP levels as well as through phosphatases are all likely to be involved in the brain's intricate regulation of its own properties. □

1. Holz, G.G., Rane, S.G. & Dunlap, K. *Nature* **319**, 670 (1986).
2. Pfaffinger, P.J., Martin, J.M. Hunter, D.D., Nathanson, N.M. & Hille, E. *Nature* **317**, 536 (1985).
3. Breitwieser, G.E. & Szabo, G. *Nature* **317**, 538 (1985).
4. Levitan, I.B. *J. memb. Biol.* **87**, 177 (1985).
5. Rane, S.G. & Dunlap, K. *Proc. natn. Acad. Sci. U.S.A.* **83**, 184, (1986).

Charles F. Stevens is in the Section of Molecular Neurobiology at Yale University School of Medicine, New Haven, Connecticut 06510, USA.

More stamps for Halley's comet

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

(Courtesy of Crown Agents stamp company Ltd, 2 Carshalton Rd, Sutton SM1 4RN, UK.)

Human energetics

Making headway in Africa

from R. McNeill Alexander

WOMEN of some African tribes carry huge loads, up to 70 per cent of their body mass, on their heads. Luo women balance the load on top of their heads, and Kikuyu women let it hang behind their backs, supported by a strap that runs across their foreheads. In either case, the feat seems remarkable. It seems even more remarkable in the light of a physiological investigation by Geoffrey Maloij and colleagues, which is reported on page 668 of this issue.

The authors measured the oxygen consumption of Luo and Kikuyu women walking on a motorized treadmill with and without loads on their heads. Loads of up to 20 per cent of body mass have no perceptible effect on oxygen consumption and, presumably, on energy cost. Loads of 70 per cent of body mass increase oxygen consumption by 50–60 per cent. Army recruits carrying backpacks perform much less well: 20 per cent loads increase

our bodies at each step, which makes muscles do work, producing potential and kinetic energy. The energy is then discarded and the process repeated. A lot of the energy that would otherwise be needed for walking is saved by shifting energy back and forth, pendulum fashion, between the kinetic and potential forms.

Maloij *et al.* suggest that a load on the head does not rise and fall or accelerate and decelerate as much as the rest of the body. If the load moves steadily along, as

if on wheels, it can, in principle, be carried without energy cost. There are no data that show whether the African women make their loads move steadily, and, in any case, steadiness would give no energetic advantage if it depended on bending and extending the back by muscle action.

It might be possible to achieve steadiness in another way, for example, by using the principle of vehicle suspension systems. To give a reasonably comfortable ride, an automobile must be mounted on springs. The wheels may bounce up and down on a rough road but the body will move much more steadily if the springs are compliant enough to bring the natural frequency of the vehicle well below the frequency of the vibrations.

Perhaps the African women use some

W.S. Stiles (1901–1985)

WALTER Stanley Stiles died on 15 December 1985. A physicist and mathematician by training, he spent most of his professional life in the Light Division of the National Physical Laboratory at Teddington, Middlesex. In his early days, great store was laid by photometry — after all, it was the division of light, not of radiation — and Stiles's life-long interest in the eye developed under the tutelage of Guild and Walsh.

Stiles was better equipped to establish rigorous criteria for the control of visual stimuli than to come to terms with the vicissitudes arising from biological variations. In the early 1930s he examined parameters of glare and, with B.H. Crawford, established the concept of veiling glare. This introduced a powerful method for monitoring the response of the eye that has stood the test of time. The precision of his thinking, coupled with his imperturbable courtesy and tact, made him a good choice for general secretary of the International Commission of Illumination; he maintained an interest in it for over 25 years.

J.W.T. Walsh, the head of the Light Division, appreciated the young man's original mind and so allotted his duties to permit him to carry out time-consuming experiments, an act which Stiles was to acknowledge handsomely even decades later. These studies dealt with the relationship between the stimulating power of light and the angle of incidence of light on the retina, particularly at the visual centre of the retina, the fovea. This is known the world over eponymously as the Stiles–Crawford effect. There had been earlier hints that the effective illumination of the retina was not proportional to the pupillary area, but the various physiological parameters — retinal location, level of adaptation and stimulus wavelength — were tackled in exemplary detail by Stiles and Crawford.

Stiles's third major investigation, in the 1950s, was of the colour mixture functions

of 'normal' observers, using a large (10°) test-field. Guild and Wright had previously produced standard data for a 2° field, and extending these observations to another operational level was important. Although the difference between the two standards was relatively small, Stiles and Burch's great effort helped to provide about two thirds of the characterization needed for a complete understanding of human chromatic performance.

In two respects, Stiles anticipated the work of others. He grappled with the quantal nature of the visual stimulus before this was done by Hecht, Schlaer and Pirenne but, as he told me, failed to see before they did how wavelength dependence was involved. Furthermore, during the war he attempted to determine a density-difference spectrum for the living human retina using fundus reflectometry. He came literally face-to-face with his subject but his visual photometry was not accurate enough to detect changes that were later easily quantified with photo-electron multipliers. None of this work was ever published but presumably remained in the notebook he used to amass under his arm whenever air-raid sirens sounded.

In his later years, Stiles continued to refine the two-colour method of the thirties (Π-mechanisms) and unremittingly helped younger followers in his footsteps. His advice was in great demand and, a few months before he died, he was delighted to be invited to and to attend a meeting in Madrid.

He has left us the legacy of precision. His earlier seminal papers, including those published just after the war, presented the data in terms of curves. "In those days", an expert in colour recently remarked acidly, "one did not publish data points." When Stiles changed tack and started publishing both data points and curves, the smoothness of the latter remained unimpaired. This cannot be said of many workers in his field.

R.A. Weale

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

R.F. Ker

A woman carrying a load in Malawi.

oxygen consumption by 13 per cent and 70 per cent loads by nearly 100 per cent.

But do not assume that you would find it more restful to carry your suitcase on your head. Experiments with people who are not accustomed to carrying loads on their heads show that the energy cost of carrying a load is the same whether the load is carried on head or back, and is much more than the cost to the African women. Further, inexperienced people cannot safely manage more than 15 per cent of their body mass on their heads.

How can African women carry loads so economically? Have they discovered some behavioural or physiological trick? Maloij and colleagues make an interesting suggestion. The energy required to move a load at constant speed over level ground is, in principle, nothing. Walking requires energy largely because we raise and lower, and accelerate and decelerate,

trick of posture to make their backs more compliant. The idea is attractive, but a crude calculation is discouraging. People usually take about two steps per second, so the back would have to be compliant enough to bring the natural frequency of the load below about 1 Hertz to reduce its vertical movements significantly. The standard formula for the frequency of spring-mass systems shows that the spring of the back would have to be soft enough to be compressed 0.25 metres by the load. Such compliance does not seem feasible, even if the back were allowed to bow out in an unusual curve. In any case, carrying a large head-load with the back bent seems dangerous. Flexible columns carrying loads are apt to collapse by buckling, and the danger is worse if the column is bent initially.

It would be interesting to attach accelerometers to the head-loads to find out how steadily they are carried. It would

also be interesting to know whether there is anything unusual about the curve of the women's backs. Meanwhile, there is something more to wonder about. Taylor *et al.* (*J. exp. Biol.* **86**, 9; 1980) have argued that most of the energy cost of running is associated with the generation of force in muscles, rather than with performance of work. Therefore, an animal carrying a load of x per cent of body mass must increase its energy consumption by x per cent. The evidence available at the time seemed to support this conclusion, but the authors have now discovered that Luo and Kikuyu women can carry 20 per cent of body mass without increasing their oxygen consumption. Is there a lesson to be learned from these women about the fundamentals of muscle energetics in locomotion? □

R. McNeill Alexander is Professor of Zoology at the University of Leeds, Leeds LS2 9JT, UK.

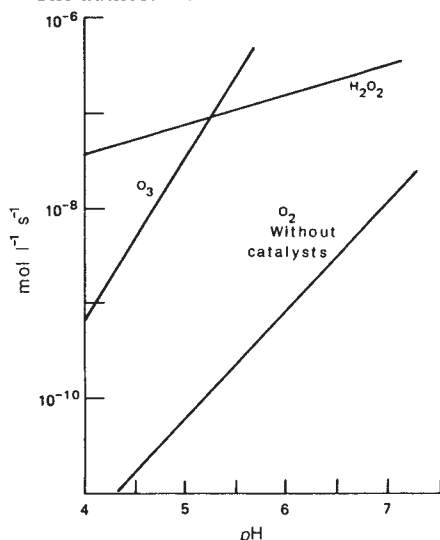
Atmospheric chemistry

Hydrogen peroxide in cloudwater

from S.A. Penkett

HYDROGEN peroxide is perhaps best known as an aid to pulchritude, although it occurs naturally and has now been shown to be a common constituent of cloudwater. Older chemistry textbooks refer to the presence of H_2O_2 in the atmosphere almost as a piece of folklore, and ascribe its presence to atmospheric electricity. In fact, a recent study using highly specific analytical methods shows that it is much more likely to be formed as a result of photochemistry of atmospheric trace gases (Kelly, T.J., Daum, P.H. & Schwartz, S.E. *J. geophys. Res.* **90**, 7861; 1985).

The authors measured cloudwater sam-



Conversion of S(IV) to S(VI). The pH dependence of the reaction rates is for the systems H_2O_2 , O_3 and O_2 .

ples collected in free air by aircraft; they found that the H_2O_2 concentration varies from less than 0.3 to 72 μM , corresponding to gas-phase concentrations up to 0.9 p.p.b.v. (parts per 10^9 by volume) with an average of 0.1 p.p.b.v. These values are roughly equivalent to a much more extensive series of measurements made on cloudwater samples collected on a mountain top by Mohen and co-workers (NATO Advanced Study Institute Conference, 1983), who show a substantial seasonal dependence of H_2O_2 , with larger values in the summer. The gas-phase concentrations estimated from cloudwater, however, are significantly less than those observed directly at the same time of year (October) at a similar location (Calvert, J.G. *et al. Nature* **317**, 27; 1985).

Interest in atmospheric H_2O_2 has increased substantially since it was realized that this substance could be the most efficient droplet-phase oxidant for SO_2 , leading to the production of SO_4^{2-} in cloud and rainwater (see figure). At cloudwater pH, which is typically less than 5, peroxide oxidizes dissolved SO_2 more rapidly than ozone or oxygen in the absence of large concentrations of metal catalysts.

Kelly *et al.* measured samples of cloudwater with specific detectors because previous attempts to estimate H_2O_2 in the gas phase had resulted in analytical artefacts (see Lazrus, A. *et al. Analyt. Chem.* **57**, 917; 1985). The new results show that the presence of gas-phase SO_2 inside clouds and cloudwater containing H_2O_2 is mutually exclusive, and that the sum of H_2O_2 and sulphate ions in the

aqueous phase is much closer to the newer, artefact-free gas-phase H_2O_2 concentration measured by Calvert's group at the same time of year (about 2 p.p.b.v.). Measured gas-phase concentrations agree closely with the values calculated by Calvert *et al.* who used a sophisticated model of atmospheric photochemistry, and moreover they often exceed measured concentrations of gaseous SO_2 . All these findings strongly implicate H_2O_2 as the predominant oxidant for SO_2 in many circumstances.

Other oxidation mechanisms for SO_2 are also possible and oxidation by hydroxyl radicals in a homogeneous gas-phase process is probably of equal efficiency to the H_2O_2 droplet process. Oxidation in droplets by free radicals collected from the gas phase (see Chameides, W. & Davis, D. *J. geophys. Res.* **87**, 4863; 1982) appears to be a rare phenomenon, although according to Kelly *et al.*'s data, so is oxidation by ozone. Both of these conclusions may be modified when further data are collected. It is quite possible that slower oxidation of SO_2 by ozone in the winter clouds may be relatively more important when the concentrations of other oxidants are low.

Recent understanding of the role of peroxides has led to the suggestion that atmospheric oxidation, and therefore acidification, may often be oxidant-limited. This has certainly complicated the issue on the best course to be adopted in addressing the acid-rain problem, as it is possible that reductions in acid deposition in remote areas may not be a linear function of emissions of gases such as SO_2 in highly populated areas. To add a further complication, many scientists now believe that oxidant damage to forests in Germany can be more significant than damage caused by acid deposition. This includes damage by H_2O_2 at the sort of levels measured in Kelly *et al.*'s cloudwater studies. It is clear we need to understand the mechanisms responsible for the production of such oxidants. □

S.A. Penkett is NERC Reader in Atmospheric Chemistry at the University of East Anglia, Norwich NR4 7TJ, UK.



Reconstruction of the head of the giant camel *Megacamelus merriami* which lived in North America about 2 million years ago. From *Smithsonian Contributions to Paleobiology* No. 57 (Washington DC; 1985), by Jessica Harrison.

Fractal harmonics and rugged materials

SIR—Fractal analysis¹ has been applied to the analysis of irregular surfaces and motions in a wide variety of engineering and scientific fields. In particular, the fractal dimension is used to characterize particle shape^{2,3}. This work has led the author to find "fractal harmonics" or equilateral polygons within closed irregular curves. While mathematically important in their own right, fractal harmonics may play a central role in future characterization of rugged materials.

Ruggedness of a closed curve can be characterized using the hand and dividers fractal technique. An arbitrary point is chosen on the curve and the dividers are swung within the limits of some protocols (clockwise–anticlockwise, swinging from inside or outside the curve) to find the next point on the curve a set distance from the starting point. Ultimately the curve is approximated by a polygon which is equilateral but for the closure side between the final point and starting point. Fractal dimension is found from the increase of polygon perimeter with decrease in step length. The greater the increase, the greater the dimension, which can assume values between 1 and 2.

If, instead of closing the final point to the starting point, the structured walk is continued to traverse the curve repeatedly, the vertices of the polygon at successive traverses converge, and the dividers start to walk in their own footsteps. Thus an equilateral polygon is set up within the curve, usually with surprising rapidity. Examples of such polygons within the outline of West Virginia are given in Fig. 1. Similar solutions often persist for small variations in step length by small changes in orientation of the polygon within the curve. For a sufficiently large variation in step length the solution changes. For example, by decreasing the step length, the polygon may change from a square to a pentagon. Moreover, some of these

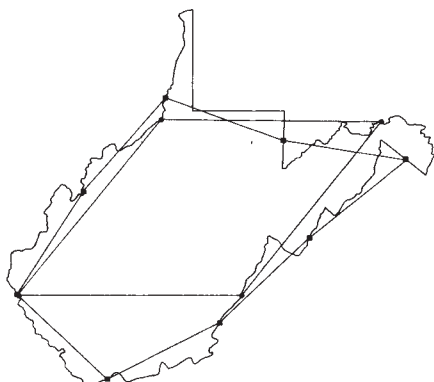


Fig. 1 Fourth and eighth fractal harmonics of the outline of West Virginia. Starting point was the northernmost point, dividers were swung from the outside in. Analysis was clockwise

changes are accompanied by more complex solutions (see Fig. 2) which generally exist over a smaller range of step lengths. Different solutions can be found by starting the analysis at different points on the curve and by traversing either clockwise or anticlockwise.

Although convergence takes longer and solutions are generally more complex with increasing regularity of the curves, I have found no curve other than the circle for which no solution exists. Analysis of the circle shows that solutions exist only when the angle subtended by the step length is equal to a rational multiple of 2π . This is

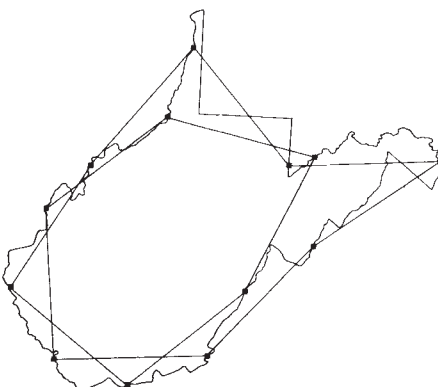


Fig. 2 The 6.5th harmonic of West Virginia, found using the same protocols as Fig. 1. The polygon has 13 sides and traverses the curve twice to achieve closure.

because re-orientation in a rotational sense permits no variation of the analysis in a constant radius curve.

The nature and stability of fractal harmonics will play an important role in future characterization of rugged particle shapes and curves. Since a change in solution is generally accompanied by a significant jump in polygon perimeter, fractal harmonics can also be used to explain error and scatter of data in performing conventional² and digital³ hand and dividers fractal analysis, where fractal dimension is deduced from a plot of perimeter versus step length.

NIGEL N. CLARK

Particle Analysis Center,
West Virginia University,
223 White Hall,
Morgantown,
West Virginia 26506, USA

1. Mandelbrot, B.B. *Fractals: Form, Chance and Dimension* (Freeman, San Francisco, 1977).
2. Kaye, B.H. *Particle Characterization* Vol. 1, 14–21 (1984).
3. Clark, N.N. *Power Technology* (in the press).

Bird dropping research continues apace

SIR—Recent speculators on the water-repellency or otherwise of bird droppings¹ seem unaware that this is of major economic importance to the power industry. High-voltage transmission lines normally consist of conductors suspended by insulators from grounded crossarms. In open

country, birds view the crossarms as convenient perches. Predictably, insulators in such areas are frequent recipients of droppings, sometimes causing power cuts via flashover of the insulators in wet weather.

The physics of this phenomenon is believed to involve low-current electrical arcs elongating over insulating surfaces^{3,4}. This process requires a continuous conducting film. If the surface is non wetting, the water forms small droplets (beads) and flashover occurs by distortion of the drops by the electric field⁵. This is unlikely in the case of power-line insulation since the electric field required is several orders of magnitude higher than typical operating fields. Thus, G.A.W. Rook is probably correct in surmising² that bird droppings render the surface more wetting, perhaps due to an ingredient in bird bile.

Further support for the hydrophilic view comes from Soviet researchers who suspended unenergized high-voltage insulators under a perch in a vulture cage. When sufficiently coated, the insulators were removed, energized and subjected to mist. Flashover occurred at voltages typical of wetted surfaces. A mildly repulsive photograph of the vulture-cage arrangement has been published⁶.

Wettability is not the only physical property to have attracted the interest of researchers. Viscosity and electrical conductivity have been closely studied. When large predatory birds defaecate near extremely high-voltage (>500 kV) insulators, the resulting viscous, electrically conducting jet can trigger sparkover by reducing the air gap. Fascinating side-issues of hydrodynamical stability are involved. Ordinarily such a jet would break up because of sausage-mode pinch instabilities caused by surface tension. When the jet is very close to the insulator, this normal capillary break-up is accelerated by electrostatic forces. Under some conditions, however, the reverse may be true, since such jets can be stabilized by longitudinal current-flow, produced perhaps here by corona at the ends of the jet⁷.

To simulate the phenomenon, engineers at the Bonneville Power Administration in the United States, after consulting with avian experts, designed a mechanical cloaca consisting of a pressure chamber with an adjustable-diameter orifice⁸. A balloon within the chamber contained raw scrambled eggs (for correct viscosity) doped with salt (for correct electrical conductivity). The doping level was determined from measurements on rehydrated cage scrapings from a local zoo. A solenoid operated needle broke the balloon on command, discharging the contents.

In full-scale tests conducted at 500 kV, the mechanical cloaca operated perfectly, resulting in spectacular electrical fire-works. As a result of this study, spikes were installed on crossarms to discourage

roosting. Animal rights activists will be pleased that no living birds were injured in the research, and that a hazard to wild birds was reduced.

DAVID C. JOLLY

PO Box 931, Brookline,
Massachusetts 02146, USA

1. Roodyn, L. *Nature* **317**, 581 (1985).
2. Rook, G.A.W. *Nature* **318**, 237 (1985).
3. Jolly, D.C. *J. Franklin Inst.* **294**, 483-500 (1972).
4. Jolly, D.C. & Chu, S.T.J. *Phys.* **50**, 6196-6199 (1979).
5. Cheng, T.C. Jolly, D.C. & King, D.J. *IEEE Trans. Electrical Insulation EI-12*, 208-213 (1977).
6. Maikopar, A.S. *Elektrichestvo* No. 2, 22-25 (1965).
7. Melcher, J.R. & Warren, E.P. *J. Fluid Mech.* **47**, 127-143 (1971).
8. West, H.J. Brown, J.E. & Kinyon, A.L. *IEEE Trans. Power Apparatus & Systems PAS-90* 1627-1630 (1971).

SIR—Not all bird droppings repel water on car roofs. In an attempt to support an insurance claim, a series of bird droppings were observed on various vehicles. It appeared in a limited series of eight sightings that only fish-eating birds' droppings have this effect: three seed-eaters, a fruit-eater and a crow had negligible repellent action. Three coastal seagulls, however, gave dramatic responses. I concluded that the cause is dietary oils: the high nitrate and acid content of most guano tends to wet rather than dry an exposed surface.

The repellent effect is seen more clearly on newer cars (post 1980) which have a polyurethane coating. Older baked paints have a less homogenous surface, and are, my insurers tell me, more liable to corrosion by any droppings.

C.A. MICHIE

University College,
Oxford OX1 4BH, UK

An alternative to archaeobacterial dogma

SIR—A recent *News and Views* report by Garrett, entitled "The uniqueness of Archaeobacteria", might more aptly have been titled, "The orthodoxy of Archaeobacteria", for it failed to report accurately the reasons that Archaeobacterial dogma is being challenged¹. The conflict at the EMBO Workshop on the Molecular Genetics of Archaeobacteria centred around two competing evolutionary trees. Each tree relates all known organisms from five major evolutionary groups. These trees are of fundamental importance because they describe the early history of all living life forms. If we are to understand molecular evolution, it is vital that we determine the correct tree.

A critical challenge facing molecular evolutionists is how to deal with both fast- and slow-clock organisms in different branches of the same tree. I believe the archaeobacterial tree is an artefact produced by greatly unequal rates in different arms of the tree. This type of situation has been analysed² and it has been found that under conditions of greatly unequal rates, fast-clock organisms will be placed with fast-clock organisms and slow-clock

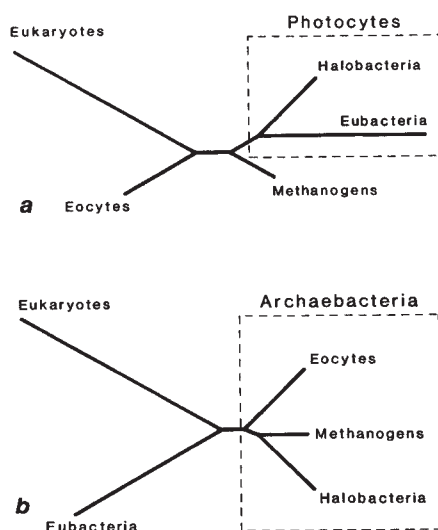


Fig. 1 *a*, The combined photocyte-eocyte unrooted evolutionary tree. *b*, The archaeobacterial unrooted tree.

ones with slow-clock organisms. The archaeobacterial tree places fast-clock organisms together, whereas our eocyte and photocyte trees do not.

Specially, we proposed the tree shown in Fig. 1*a*, based on our analyses of three-dimensional ribosome structure. Our tree makes two unique predictions that have important consequences. First, our data indicate that the eukaryotes are evolutionarily closest neighbours to the eocytes³. This would then explain the many eukaryotic properties of the eocytes, including the recent observations reported at Martinsried, that eukaryotes and eocytes both have similarly constructed ribosomal operons, in contrast to the common pattern which is found in eubacteria, methanogens and in halobacteria.

Our second proposal (shown at the right side of our tree), is that the eubacteria and halobacteria are evolutionarily closest neighbours⁴. Again, this was based on the three-dimensional structure of ribosomal subunits from the major groups. It also brings together all the photosynthetic prokaryotes, the halobacteria and the eubacteria, into a single group. As a result we called the larger group the photocytes and suggested that photosynthesis could have been a primitive property of the group.

The alternative archaeobacterial tree proposed by Woese and coworkers is shown in Fig. 1*b*. According to this tree, derived from sequence data, the slow-clock organisms on the right side of the tree are classified as the archaeobacteria. The two fast-clock organisms, indicated by long arms of the tree, are both on the left side of the tree. This is precisely the effect expected when evolutionary rates differ greatly in different branches of a tree. Hence, primary sequence data, as *presently analysed*, neither support nor disprove the archaeobacterial tree or any other tree.

Our tree, based on ribosome structure, appears to be insensitive to this effect, probably because three-dimensional structures are much more conserved than are primary sequence data. The best documented examples of this are provided by protein structures, determined by X-ray crystallography, that are well conserved between distant organisms and yet have primary sequences that have diverged so much that they are scarcely recognizable. Under conditions of lower absolute evolutionary rates it is known that trees will reflect the correct topology. Hence, three-dimensional structures, because they are better conserved than primary sequences, seem to be ideally suited for determining the deepest branching in the evolutionary tree.

Molecular evolution is on the way to establishing an accurate classification of organisms. This classification must be built on phylogenetically correct trees. I believe that the optimistic message that came from Martinsried was "let's wait and see which tree is correct". Time will reveal the correct tree and with it will come the correct classification at the kingdom level.

JAMES A. LAKE

Molecular Biology Institute,
and Department of Biology,
University of California,
Los Angeles, California 90024, USA

1. Garrett, R. *Nature* **318**, 233 (1985).
2. Felsenstein, J. *Syst. Zool.* **27**, 401 (1978).
3. Lake, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 3786-3790 (1984).
4. Lake, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 3716-3720 (1985).

Animal model for Epstein-Barr lymphoma

SIR—I was surprised to read in the *News and Views* section of the issue of 21 November 1985, an erroneous statement. A.J. Beale, in his piece "Epstein-Barr virus: Dream or reality of a vaccine" states that "Epstein *et al.* have developed an animal model of EBV-induced lymphomas in cotton-top tamarins (*Saguinus oedipus oedipus*)"¹.

In fact, Dr Thomas Shope and I and other colleagues developed this experimental model of lymphomagenesis by EBV in the early 1970s. It depended on the production of the high titted immortalizing B95-8 strains of EBV² and showing that the cotton-top marmoset was susceptible to induction of malignant lymphoma on inoculation with B95-8 virus³.

GEORGE MILLER

Department of Pediatrics,
Yale University
School of Medicine,
PO Box 333, New Haven,
Connecticut 06510-8064, USA

1. Beale, A.J. *Nature* **318**, 230 (1985).
2. Miller, G. & Lipman, M. *Proc. natn. Acad. Sci. U.S.A.* **70**, 190-194 (1973).
3. Shope, T.E., Dechairo, D. & Miller, G. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2487-2491 (1973).

Angles on intelligence

H. J. Eysenck

Handbook of Intelligence: Theories, Measurements and Applications.

Edited by Benjamin B. Wolman. Wiley:1985. Pp.985. \$64.95, £66.45.

THE concept of intelligence has a paradoxical standing in psychology. On the one hand it has always been regarded as the prime example of the possibility of measuring mental qualities with considerable reliability and validity; on the other, it has been subjected to a good deal of criticism and indeed obloquy. There have been unending arguments about many different aspects of intelligence, from the generality of the concept to the question of the relative importance of genetic and environmental factors. Much of the controversy has been uninformative, probably because there are at least three different concepts of "intelligence" which are frequently mixed up by the various parties involved.

In the first place we have what Donald Hebb called "Intelligence A" (that is, the biological foundation for all cognitive behaviour). Then we have "Intelligence B" (the application of this intelligence to life problems, which is much influenced by personality, education, cultural factors and many other extraneous influences). Last, we have "Intelligence C" (psychometric intelligence as measured by IQ tests). To have a worthwhile scientific argument, we must specify which of these we are talking about, and unfortunately the many authors in this book often fail to be clear about the particular meaning of intelligence they intend. This accounts for at least some of the contradictions which are apparent in their many diversified contributions. An introductory chapter discussing the concept of intelligence from the historical and philosophical points of view might have helped to clarify matters; unfortunately the book jumps in at the deep end, and never attempts seriously to unravel these tangled issues.

Actually the recent appearance of not one, but two handbooks on intelligence (the *Handbook of Human Intelligence*, edited by Robert Sternberg, was published by Cambridge University Press in 1982) indicates a renewed interest in the subject after several years of relative neglect. Both handbooks cover the new developments although they stress rather different aspects. There can be no doubt that this rebirth of interest is welcome; intelligence is a central concept in psychology, and the discovery of anything new in this field must inevitably lead to a re-evaluation of its meaning.

As both books show, there is much agreement on two issues. One is the im-

portance of genetic factors; this is not seriously doubted by any expert, although outsiders such as Kamin are still willing to argue the toss on the basis of ideological rather than scientific notions. Second, agreement on the importance of a general factor of intelligence (*g*) is almost but not quite unanimous; here too the aberrant notions of the 1960s and 1970s have had to bow to the strength of the evidence. The chapter in *Handbook of Intelligence* by J.P. Guilford on his structure-of-intellect model still stoutly argues the case against *g*, but his methods of analysis have been sharply criticized by almost all experts in the field, including John Horn and Nathan Brody, both of whom also contribute chapters to this book. Guilford's belief in the existence of some 120 "intelligences", all independent of each other, is simply not tenable in view of the fact that all the relevant tests are highly intercorrelated, and also correlate highly with typical IQ tests. Guilford argues somewhat ingeniously that some of the observed correlations are quite low, but fails to mention either that most of them are very high, which should be impossible on his theory, or that those tests which have low correlations are practically always those which have very poor reliabilities.

Guilford's chapter typifies one of the weaknesses of the present book, namely a tendency to let the authors ride their particular hobby horses. The worst example is probably a chapter in which Allen and Nadeen Kaufman blow the trumpet for the Kaufman Assessment Battery for children, praising their own work and neglecting to mention the many criticisms that can be, and have been made of their approach. What, for instance, might the validity of this test be? Practically the only information given is that it correlates around 0.7 with the Wechsler IQ, that is, about the same or slightly less than correlations usually observed between different IQ tests. Actually the correlation is about the same as that which Nathan Brody suggests, from a survey of the evidence on different choice reaction times, can be obtained between IQ and reaction time! One would need a great deal of persuasion before accepting the claims made by the Kaufmans for their new test, and it is certainly unusual to allow protagonists the chance to advertise their wares in a book such as this.

There is no question that many of the chapters are very informative, particularly

those on genetics, neurological foundations, the life-span development perspective, the validity of tests of intelligence, environment and IQ, and several others. Nevertheless there are other fundamental faults. The book is parochial both in its authorship (almost exclusively North American) and in its inclusiveness. In the chapters dealing with factor analysis, there is no mention of such figures as Meili in Switzerland or Jaeger in West Germany. Equally, there is no mention of Lienert's important contribution — his demonstration that the number of factors extracted from a matrix of intercorrelations between IQ tests is significantly smaller for children who are emotionally unstable as compared with those who are emotionally stable — or of several other German studies replicating and extending this finding.

In the chapter on genetics, too, no place has been found for discussion of such work as that of Volkmar Weiss in East Germany, or of Lippovechaja, Kantonistova and Chamaganova in the USSR. It is not only that these authors have shown that genetic factors are equally important in socialist countries as they have proved to be in capitalist ones, a vital point in itself; Weiss has also made important theoretical contributions to the analysis of genetic factors in intelligence which run counter to much current thinking.

Of the many omissions, the most striking is probably that of a chapter dealing with the speed-of-information-processing theory of intelligence, which has found much support in recent work on the correlation between reaction time and intelligence, and also the relationships between IQ and evoked potentials. There is no discussion of this research in the chapter where it should belong, namely that on neurological foundations of intelligence; Nathan Brody is the only one, for instance, who mentions the work of the Hendrickson's and he does so only in passing, and without mentioning the equally important work of Schafer, Robinson and others. This is fair enough in a chapter dealing with the validity of tests of intelligence, which has rather different concerns; but to include a chapter on a theory such as Guilford's, which is almost universally rejected, and not include one on these modern theories which have a great deal of experimental support, and seem to change our interpretation of intelligence test results quite drastically, is inexcusable. It is curious that Sternberg's *Handbook* erred in exactly the same direction; is it possible to detect the hidden hand of the *Zeitgeist* here?

A last few words may be appropriate, comparing the two books. Sternberg is stronger on the theoretical side, Wolman on the applied. Wolman has a whole section on "applications", dealing with men-

tal health, clinical applications, clinical uses of the Wechsler, and educational applications, as well as chapters in previous sections on group tests of intelligence, the assessment of mentally retarded individuals, giftedness, transcultural assessment, and clinical assessment of children's intelligence. Factor analysis is much more prominent in Wolman than Sternberg, but the latter is far more concerned with cognitive theories, and particularly various componential theories; these are not even mentioned in the subject index of Wolman. From the point of view of the buyer, the Sternberg book is more academic, the Wolman more applied — as is not surprising, in view of the fact that it is dedicated to the memory of David Wechsler, whereas the Sternberg book deals more with the kind of theories associated with the name of the editor. It

is impossible to say that one book is better than the other; both make important contributions, and many of the chapters in both are excellent. In addition both offer a useful collection of references.

Neither, however, really lives up to the claims one would normally make for a handbook. A handbook that really dealt with the important issues that arise in the measurement of intelligence, did not leave out non-English-speaking authors who have made seminal contributions to the subject and did not evade discussion of important recent studies would be well worth having. Until such a volume is produced, we will have to make do with what we are offered. □

H.J. Eysenck is Emeritus Professor at the Institute of Psychiatry, University of London, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.

The Faraday effects

Michael Shortland

Faraday Rediscovered: Essays on the Life and Work of Michael Faraday, 1791–1867. Edited by David Gooding and Frank A.J.L. James. *Macmillan, London/Stockton, New York: 1985. Pp.258. £30, \$60.*

"SCIENTISTS", remark the editors of this volume, "are inveterate myth builders". If this is so — and who could survive without folktales and fictions? — few could have been so triumphantly romanced as Michael Faraday (1791–1867). To Maxwell, he was a friend "entirely free from pride and undue self-assertion"; to Tyndall, a conquering hero "who took no cities but captivated all hearts"; and throughout Bence Jones's reverential *Life and Letters of Faraday* (1870) a paragon of "supernatural" virtues. Only the occasional discordant note could be heard, and that invariably from France, a land with its own fabled *savants*. Which shows perhaps the universal need for scientific heroes, and the fact that myths do not on the whole travel well.

For a nation which professed to have abandoned idolatry Victorian England offered itself remarkably quickly at the shrine of science. Biologists, geologists, chemists and physicists having become a new scientific clerisy, it followed that they should be accorded respect and approached with veneration. Remembering an audience she had with Faraday in 1850, Mrs Cornelia Crosse spoke of the "awe" which overcame her:

With the knowledge that we were approaching the Arcana of Science, I was in no condition of sympathy with the fools who rush in, but felt restrained by the reverent attitude of those who fear to tread on sacred ground.

Such devotion could not have been won without Faraday's consent. Indeed, one of the pleasures of this set of original essays is that it forsakes the easy opportunity to ridicule this particular "Faraday effect" in favour of the more difficult task of unravelling its roots, growth and impact. Stepping beyond the enormous archive of documentation left us by Faraday, the contributors begin to take a just measure of the man. We can then see how little is revealed — and how much concealed — by invoking as explanations for Faraday's achievement those favourites of the nineteenth-century biographer, "native genius", "self help" and "moral rectitude". Instead comes the attempt, partial but successful, to place Faraday's life and work in context.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Plain no more — Michael Faraday, with his wife, c. 1840.

Many new themes emerge from this enterprise. By comparison and contrast with Wheatstone, Ampère and Davy we can begin to follow the development of Faraday's interests and the impact of his theoretical and practical innovations in the domains of electrodynamics and electrical engineering. But to account for his success as a scientist we need to understand how he suited that role and how he made that role suit him. More than any equivalent figure, Faraday's private life and public image were intertwined. For 32 years he was Fullerian Professor of Chemistry at the Royal Institution, and with that position came an apartment in the building. Transitions from lodging to laboratory, thence to lecture theatre, were easy and frequent. Who, then, was the real Faraday? The homely, devout Christian; the tireless, inventive experimenter; or the irresistibly eloquent public speaker? The answer is that Faraday was all and none of these. His science cannot be separated from his religious beliefs, as Cantor shows here in an exploration of Faraday's Sandemanianism. Nor can the man be disassociated from the myth: through the many photographic portraits of him, Faraday seemed to grow into his public image.

At a more intricate level, essays by Gooding and James reveal how Faraday used the lecture theatre to present to his audience as obvious, collectively-witnessed "facts" those findings which he had single-handedly, and often with immense difficulty, produced in the laboratory. Experiments served an obvious educative role, but they also reaffirmed the status of scientific knowledge as objective and reproducible. It has often seemed as though Faraday's objections to the use of mathematical operations were based on prejudice, particularly since they were expressed in an era which witnessed the ever-increasing incursion of mathematics

into large areas of physics. This volume shows, on the contrary, how integral to his style, and scientific and religious beliefs, these objections were.

No single figure emerges from this book which could replace the one we have lost. Faraday has fallen from his pedestal but acquired other dimensions and depths. It is doubtful whether a cogent identity will — or should — result from further historical researches into Faraday's life and work. One of his last requests was to remain "plain Michael Faraday to the last", but Faraday rediscovered can be plain no more. □

Michael Shortland is in the Department for External Studies, University of Oxford, Rewley House, 1 Wellington Square, Oxford OX1 2JA, UK.

Powerful techniques at work

Indu Parikh

Affinity Chromatography: A Practical Approach. Edited by P.D.G. Dean, W.S. Johnson and F.A. Middle. IRL Press: 1985. Pp.215. Pbk £11, \$20.

Affinity Chromatography: Template Chromatography of Nucleic Acids and Proteins. By Herbert Schott. Dekker: 1985. Pp.234. \$49.50.

AFFINITY chromatography has played a key part in the development of molecular biology and biotechnology. Since the technique was described, the number of applications in the biological sciences has grown dramatically. Of the four books and four proceedings of international symposia published to date which deal with the subject, the volume edited by Dean, Johnson and Middle stands out as a uniquely practical guide.

Broadly, the editors have divided their material into three parts: matrix preparation, operational methodologies and quantitative analysis. The section on matrix preparation, which amounts to almost 15 per cent of the text, is rather cluttered with details a biochemist rarely needs. The parts of the book dealing with methodology and quantitative aspects are well written and well organized. A good cross-section of examples and experimental strategies is presented, with sufficient details to illustrate various activation procedures and cross-linking agents for coupling spacers to matrices. Scores of examples of ligand immobilization with superb recipes are scattered throughout the book.

Although a brave attempt has been made to provide adequate bibliographies with original citations, many sections of the book fall down in this regard. But, overall, the editors have successfully brought together a variety of "practical approaches" and produced a most useful and handy package.

Herbert Schott's book is rather different. It is not a manual of recipes, and is limited to nucleic acid research, but is remarkable for its breadth of approach and depth of information. With the upsurge of interest in nucleic acids in general and

nucleic-acid-protein interactions in particular, the appearance of a book dealing specifically with template chromatography of nucleic acids and proteins is timely indeed.

Schott has organized the book in a logical fashion with the most basic system first. Following chapters, in which the procedures and applications become more advanced, build successively on the preceding material. Following an account of immobilization and the use of simple nucleosides and nucleotides, Chapters 2, 3 and 4 describe immobilized oligonucleotides, polynucleotides and nucleic acids. Chapters 5 to 8 deal with the chromatography of DNA and RNA (specifically messenger RNA). The last three chapters, which describe DNA-protein

interactions, are especially good.

Of particular interest to many people is likely to be the isolation of mRNA. Schott devotes an appropriate amount of space to this topic and includes a discussion of the pros and cons of the most commonly used ligands in the isolation of mRNA. The extensive table for specific mRNA, giving sources, how it was purified, and literature references, is particularly helpful.

Finally, an extensive bibliography and an elaborate index are provided. This is a book that will be an invaluable asset to all laboratories engaged in nucleic acid research. □

Indu Parikh is Associate Head of the Department of Molecular Biology, Wellcome Research Laboratories, Research Triangle Park, North Carolina 27709, USA.

Living in the past

A. Hallam

The Burgess Shale. By Harry B. Whittington. Yale University Press: 1985. Pp.151. \$21, £21.

LATE in the first decade of this century the former director of the US Geological Survey, Charles D. Walcott, crowned many years of assiduous searching for Cambrian fossils by making one of the most outstanding fossil discoveries ever. In the Burgess Pass in the southern Canadian Rockies, Walcott and his party found not only the usual types of Cambrian fossils with hard-part preservation, such as trilobites, brachiopods and sponges, but a vast variety of well-preserved soft-bodied organisms whose existence had hitherto been unsuspected. Several seasons of collecting yielded an immense number of specimens which were housed in the US National Museum in Washington. Until the late 1960s there they languished, gathering dust and largely undescribed.

Professor Whittington's book on the Burgess Shale fauna summarizes nearly two decades of his concentrated and meticulously thorough research, and that of his students and associates, which has opened a unique window on life on the sea bed early in metazoan history. The soft-bodied or poorly skeletonized fossils, in which there is most interest and which far outnumber those with hard parts, include a large variety of arthropods and worms together with primitive coelenterates and chordates. Not surprisingly, there are some fossils so unlike living organisms that they defy classification. None is more bizarre than the aptly-named *Hallucigenia*, which apparently traversed the sea bed on stilt-like pairs of legs, grasping food by means of a number of tentacles.

The exceptional preservation is the result of an unusual geological event. The

organisms lived on or within the muddy sea bed close to the foot of a submarine cliff formed by a calcareous biogenic reef. A slumping disturbance caused them to be transported suddenly into a deeper, anoxic environment beyond the reach of scavengers, and there the fossils lay undisturbed, exquisitely preserved in shale, until their discovery this century.

As with all scientific advances, as many problems have been posed as solved by research on the Burgess Shale; Whittington's work emphasizes how little we really know about early metazoan life, with its diversity of evolutionary experiment. There is no hint, for example, from any older deposits of how the extensive diversity exhibited by the Burgess Shale fauna developed through time. Why did hard parts appear in certain, but not all, animal groups in the early Cambrian? Could the development of hard parts, valuable for protection and muscle support, be the result of changes in the environment, for example the composition of sea water, or could it perhaps have depended on the evolution of collagen, which is essential for linking muscles to shells? The evolution of some groups such as the brachiopods must have been dependent upon the formation of a hard external skeleton.

Before the publication of this clearly written and well illustrated book, those of us interested in the Burgess Shale fauna were dependent upon either a few brief popular accounts or lengthy technical reports in the specialized literature. Only the reports do full justice to the skill and painstaking efforts that have been called upon to collect, prepare, describe and interpret functionally the various peculiar fossils, but we are indebted to Professor Whittington for now making the research findings of his group so much more accessible. □

A. Hallam is Lapworth Professor of Geology, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.

Textbooks supplement

Next week's issue of *Nature* includes the annual textbooks supplement, with reviews of some 90 textbooks published over the past year.

Among the books covered are revised editions of the texts by E.R. Kandel and J.H. Schwartz, A. Jolly, B. Lewin, R.W. Old and S.B. Primrose, J. March, P. Atkins and K.E. Bullen, and new publications by M.E. Begon, J.L. Harper and P. Townsend, S. Stanley, A. McBirney, J.R.L. Allen and J.R. Waldram.

Haldane as guru

N.W. Pirie

Haldane: The Life and Work of J.B.S. Haldane with Special Reference to India. By Krishna R. Dronamraju. *Aberdeen University Press*:1985. Pp.211. £14.90, \$20.50.

A "Haldane Industry", similar to the Darwin, Galileo and Newton Industries, is not likely to arise. This is partly because Haldane, though very influential, did not initiate a new way of looking at nature which was, in essence, comprehensible to all scientists, and partly because he wrote

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Haldane at the Indian Statistical Institute — " . . . too original a man to fit tidily into any political, religious or social system".

amply about himself and left few obscure corners. Nevertheless, he had so many facets of character, and contributed to so many aspects of science, that biographies of him should be written from several points of view.

Dronamraju worked with Haldane, first as a PhD student and then as a colleague, in India until Haldane's death in 1964. Although reverence and respect are intimately mixed in his account, they do not disguise the fact that Haldane often managed the practical business of running a research institute clumsily. Dronamraju deals tactfully with the reasons for the break with the Indian Statistical Institute, and then with the research unit set up in Calcutta by the Indian Council of Scientific and Industrial Research, before the final move to a less troubled existence in Bhubaneswar. The account adds little to what those who knew Haldane already knew: gossip addicts will have hoped for more.

Brief descriptions are given of Haldane's research in Britain and in India, of his serious general articles, and of his hundreds of popular articles. An important point to emerge is that payment for these articles enabled him to pay the expenses of

his colleagues to attend conferences. Also, while book reviews are usually thought of as ephemera, Dronamraju lists 69 from the *Journal of Genetics* and quotes from several of them; to judge from the samples, it seems that they would repay study as a running commentary on changing attitudes in genetics and related sciences.

Haldane went to India in 1957 for political and climatic reasons. It is clear from this book, and from other publications, that the move was in no sense a retirement, but was the beginning of a new phase of scientific work. I feel that Dronamraju overstates the extent to which Haldane accepted Hinduism. He was, of course, sympathetic to the principle of non-violence in spite of having written, many years before moving to India, that he enjoyed killing people in the First World War. But non-violence underlies almost all ethical systems — in Hinduism it may be nearer the surface.

Haldane was too original a man to fit

tidily into any political, religious or social system. The attempt to appropriate him is one point on which this book can be criticized. There are others. The text is almost unbelievably repetitive. One wonders whether the author, having shoved his notes together, ever re-read the whole. Space is wasted on what seem like excerpts from travel agents' leaflets about places visited, and on potted biographies of Haldane's more distinguished visitors. That space could more usefully have contained explanations of the nature of research projects other than Dronamraju's own.

In spite of its defects, this is a book from which anyone interested in Haldane will learn something. There are new facts in it, and pungent quotations from relatively inaccessible publications. □

N.W. Pirie is at Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JQ, UK, and is a Visiting Professor in the Leaf Protein Unit at the Indian Statistical Institute, Calcutta.

Heavenly movement

Victor Szebehely

Celestial Mechanics: A Computational Guide for the Practitioner. By L.G. Taff. *Wiley*:1985. Pp.542. \$45, £46.55.

DID not Newton do everything in celestial mechanics? Or, if not Newton, then Laplace and Lagrange? The resounding "no" to these questions given by the *cognoscenzi* is strongly supported by Taff's new book. The subject — which deals with the motion of natural and artificial bodies under the influence of gravitational forces — is very much alive today. The possible applications are legion, ranging from planetary theory to artificial satellites to stellar dynamics, and new and original contributions continue to appear in the literature.

Taff's book is an unconventional contribution to the science (or, as the author calls it, the "art") of celestial mechanics. Usually we write for our colleagues, even when we write textbooks. Here is a delightful exception, written for graduate and senior-level audiences in an informal style and including helpful analytical "tricks". Taff's view is that it is more important for the reader to see the logic of a procedure than to be intimately familiar with the details. Such a straightforward approach contrasts, for example, with the sometimes overly sophisticated presentations of the gravitational problem of N-bodies. On this point, Taff brooks no nonsense: "The astronomical community must face the fact that our understanding of the general N-body problem ends at N=2". Those of us who have dedicated

our careers to $N \geq 3$ can only — if with some reluctance — agree with him.

These remarks should not be taken to imply that the book is in any way lightweight. While it provides all the necessary background it is fully up to date and, at points, especially when dealing with Taff's own interests, will serve as much as a work of reference as a textbook. Readers will need to be familiar with the basic techniques of vector analysis, series solutions of differential equations, numerical integration methods and least squares, and with the fundamentals of classical mechanics.

There is room for improvement, but the many, mostly insignificant, errors can be easily remedied in a second edition. And, to be fair, celestial mechanics now covers so many complex subjects that nobody can claim universal expertise; is it reasonable to criticize the error in Fig.30 (which was borrowed from Moulton's book) when the chapter on the three-body problem, containing this figure, is sound, just because the reviewer is a fussy budget on this particular topic? The subject index is rather short and might also benefit from attention, and an author index would be helpful. The list of astronomical constants could be enlarged and several of the "Problems" could be made more approachable by offering additional information. But these criticisms are dwarfed by the success of an author who has shown once again that Newton, Laplace and Lagrange have left much work for us to do in this beautiful and fascinating area of theoretical astronomy. □

Victor Szebehely is R.B. Curran Professor in the Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, Texas 78712, USA.

From *J.B.S.* by Ronald Clark (OUP, 1984).

Electron microscopy of frozen-hydrated biological material

Murray Stewart & Guy Vigers

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The electron microscopy of frozen-hydrated specimens has the potential to allow biological material to be examined in conditions very similar to the native state, and so has excited considerable interest. Although not all of the problems associated with the method have been solved, the technique has already provided important insights into macromolecular structure and assembly and promises to be a powerful tool for investigating the structure of cellular components and their interactions.

ELECTRON microscopy has been invaluable in the study of the structure of biological macromolecules, their interactions and their arrangements in cells. However, traditional methods of sample preparation are rather harsh, because they aim to remove the water present in the specimen before placing it in the high vacuum environment of the electron microscope. Typically, specimens are embedded in plastic, heavy-metal salts or glucose, or replicated in platinum before they are examined in the microscope. Biologists have therefore sought methods that would enable specimens to be examined in the microscope in an environment more closely related to their native milieu¹. Examination in water would be ideal, but attempts to use wet samples² have often been frustrated by technical difficulties. A close approximation would be to examine frozen-hydrated material, and recent developments have shown that this is indeed feasible. Many of the restrictions of traditional methods can be avoided in this way and, in addition, there are exciting possibilities for exploring the conformational changes and interactions between particles in their liquid environment, an area hitherto inaccessible to investigation at high resolution by electron microscopy³. Here, we review some results obtained with frozen-hydrated specimens and discuss some of the difficulties encountered, before considering the likely impact of this technique on structural biology at both molecular and cellular levels and identifying the types of specimens that appear to be most promising.

Although the examination of frozen-hydrated biological specimens by electron microscopy had been proposed earlier⁴, it was not until 1974 that Taylor and Glaeser⁵ reported the first successful high-resolution study. By electron diffraction, they demonstrated that crystals of catalase can be preserved in ice to a resolution of 0.34 nm, although images of this material contained data to a resolution of only 1.15 nm because of movement of the stage on which the specimen was mounted^{6,7}. These remarkable images demonstrated that the intrinsic density difference between protein and ice could produce contrast in the absence of heavy-metal stains and, as protein is denser than ice, it appeared dark, giving reversed contrast compared with its appearance in negative stain. Similarly prepared bacterial surface layers showed enough contrast to enable fine structural features such as the bilayer profile of membranes to be seen⁸. Nonetheless, despite this spectacular advance, it was difficult to make suitable specimens. Furthermore, the ice in which specimens were embedded was crystalline, which, it was feared, could distort delicate particles, or force material to the surface of the ice film containing the specimen where it could be freeze-dried^{9,10}. Also, because the film was thinned by evaporation, this method was restricted essentially to samples in distilled water. It was thus a significant advance when Dubochet and colleagues^{11,12}, using simple but elegant apparatus, demonstrated that thin aqueous films could be frozen to produce a vitreous (glass-like) state¹³. This work transformed the method

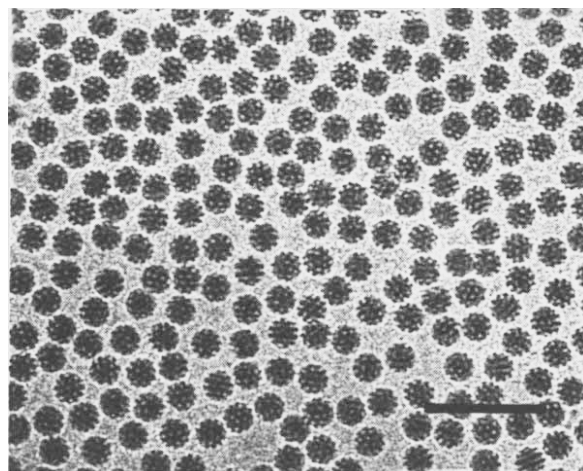


Fig. 1 Semliki Forest virus particles in vitreous ice. The icosahedral arrangement of subunits is clearly visible. (Unpublished micrograph courtesy of Dr J. Dubochet.) Scale bar, 250 nm.

so that it could be generally applied to biological specimens. These workers produced thin (100–200 nm) films containing a range of biological specimens in vitreous ice which appeared to contain no crystalline ice when examined by electron diffraction. Furthermore, specimens embedded in thin vitreous ice seemed well preserved and to retain their structural integrity to a high degree. A wide range of frozen-hydrated biological samples, prepared using these^{11,12} and other^{10,14,15} methods, have now been examined in several laboratories, and since many specimens have also been well characterized by conventional electron microscopy and other methods, the technique can be evaluated.

Virus particles

Viruses were among the earliest particles studied in vitreous ice^{9,16}. Frozen suspensions (Fig. 1) contained particles with dimensions close to those determined by X-ray diffraction methods. The symmetry of icosahedral and helical particles was well preserved and fine structural features, such as adenovirus spikes, could often be seen¹⁶. A three-dimensional reconstruction of bacteriophage T4 tail¹⁷ resembled closely the structure reconstructed from negatively stained material¹⁸ and seemed easier to interpret because problems associated with stain exclusion, uneven staining or positive staining were avoided. Another advantage was that most particles in the field seemed well preserved, whereas generally fewer well-preserved particles are seen in negatively stained specimens. This observation seems to be widespread and is important because it can make finding views of particles in key orientations easier. Tobacco mosaic

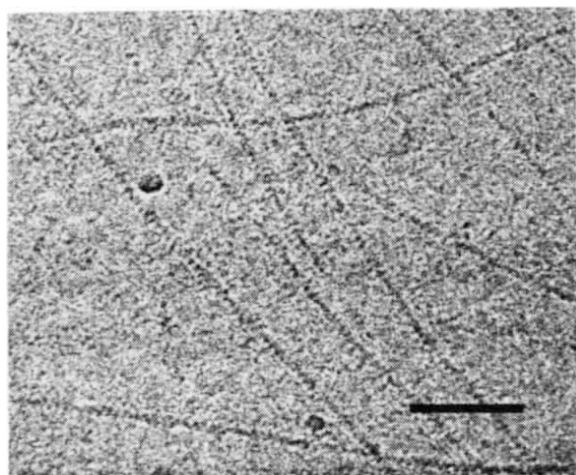


Fig. 2 Filaments of F-actin in vitreous ice. The two long-pitch helical strands of the structure can be discerned and the individual subunits are remarkably clear. (Unpublished micrograph courtesy of Dr J. Trinick.) Scale bar, 100 nm.

virus (TMV) in vitreous ice gave electron-diffraction data extending to better than 0.3 nm resolution, although images only showed data to 1.15 nm resolution in rafts of viruses and 2.3 nm in isolated particles¹⁹. In retrospect, one might argue that, since objects such as small viruses are usually well preserved in negative stain, they would appear similar in structure when viewed in ice. But superior preservation has also been obtained for some less robust particles. For example, T4 polyheads are better preserved because their cylindrical shape is not compressed, as it is in negatively stained or freeze-dried preparations¹⁹. Similarly, influenza viruses are distorted by negative stain but are spherical in vitreous ice²¹ and details of surface spikes, internal membranes and nucleic acid distribution are clearly seen. Also, as DNA is denser than protein, it can be visualized directly by adding solutes such as metrizamide to contrast-match the viral protein and solvent. When T4 phage was examined in this way, DNA in the head was clearly visible⁹ and a regular pattern of loops or rings could be seen¹⁶, corresponding to the generally circumferential arrangement of DNA detected by X-ray diffraction²². Attempts to locate the RNA in TMV¹⁹ were less successful because of difficulties caused by phase contrast (see below).

Cytoskeletal components

Helical cytoskeletal structures embedded in ice have been investigated in some detail and several aspects of their structure clarified. Fourier transforms of electron micrographs of microtubules in vitreous ice closely resembled X-ray diffraction patterns from oriented hydrated bundles²³, indicating that the structure is well preserved during freezing, even though low temperature causes depolymerization of microtubules *in vitro*. Frozen-hydrated microtubules were not flattened as they are when negatively stained and so could be analysed more easily and more reliable estimates made of their diameter, because stain was not concentrated at the edge²³. Frozen-hydrated material also showed a previously unrecognized supertwist in the helix with a pitch of several micrometres, and the rapidity of freezing enabled the mechanism of microtubule depolymerization to be explored indicating that, *in vitro*, monomers were probably removed from the body of the microtubule as well as from the ends²³. Actin filaments in vitreous ice²⁴ showed a higher proportion, compared with negative staining, of images giving diffraction patterns resembling X-ray diffraction patterns from actin filaments in whole muscle fibres. Frozen-hydrated material (Fig. 2) was used to estimate the diameter of actin filaments, which seemed greater than indicated by negative staining²⁴. Actin paracrystals in ice gave optical diffraction patterns similar

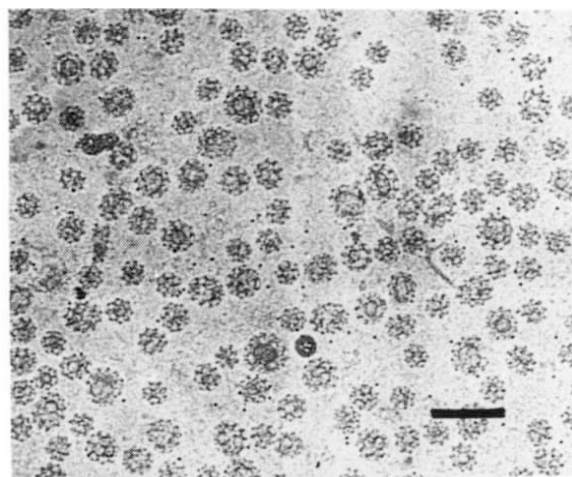


Fig. 3 Coated vesicles in vitreous ice showing both the clathrin coats and the membrane vesicles within them. (Unpublished micrograph by Dr B. M. F. Pearse and G.V.) Scale bar, 25 nm.

to those from negatively stained material (J. Dubochet and M. S., unpublished).

Electron micrographs of frozen-hydrated magnesium paracrystals of tropomyosin²⁵ showed details of the overlap between consecutive molecular ends that had been unobtainable directly from stained material²⁶. Furthermore, a single mercury label attached to a thiol residue could be detected in the presence of the protein signal in frozen-hydrated material²⁵, which enabled its location in the structure to be established directly. Previous work, using critically point dried unstained material²⁷, relied on relating the label to the protein structure by an indirect method, because the signal from the mercury was too low to be detected in negatively stained specimens, whereas the protein signal from the dehydrated unstained material was too weak to establish molecular positions.

Cytoplasmic structures, especially those that are open or loosely attached, are sometimes difficult to preserve using conventional methods; preservation in ice can therefore offer the chance to obtain substantial new information. For example, ice prevents the collapse of lipid vesicles that usually occurs when specimens are negatively stained^{20,28}. Similarly, coated vesicles are fragile structures that collapse in negative stain but are beautifully preserved in vitreous ice (Fig. 3). Both the spherical membrane vesicle that underlies the coat and the tracery of clathrin triskelions that make up the outer cage (within which the vesicle is suspended) are clearly visible.

Crystalline sheets

Although arrays of membrane proteins have been examined extensively in negative stain or glucose, frozen-hydrated specimens often produce important new insights into structure and function because the whole particle, rather than only those portions outside the membrane, can be examined. Gap junctions have been investigated in considerable detail²⁹ and frozen-hydrated specimens (Fig. 4) were used to examine the structure in two calcium-sensitive states that had been characterized previously by X-ray diffraction³⁰. Three-dimensional reconstructions to a resolution of 2 nm were obtained from material in crystalline ice and showed the shape of the entire subunit, whereas previous work using negative stain only showed those portions of the structure located outside the membrane. There was good agreement between the extra-membranous regions common to both reconstructions, but the continuity of the subunits through the membrane made interpretation of the calcium-induced structural changes more straightforward. Thus, the reconstructions indicated that the switching between states occurs by a small cooperative rearrangement involving the tilt

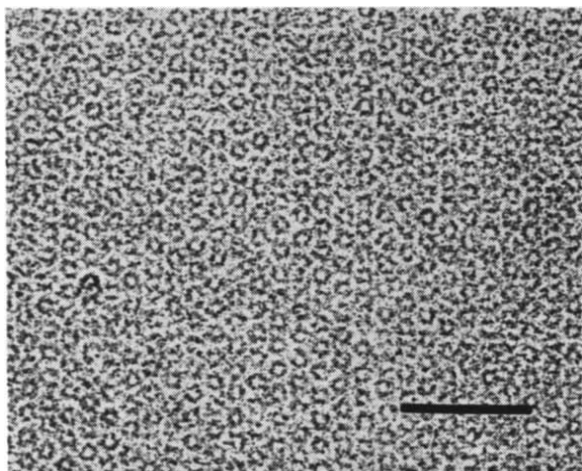


Fig. 4 Two-dimensional array of gap junctions in crystalline ice showing a hexagonal lattice of annular units. (Unpublished micrograph courtesy of Drs E. Gogol and P. N. T. Unwin.) Scale bar, 25 nm.

of the subunits, which could account for the effect of calcium on gap-junction permeability *in vivo*²⁹. Similar methods were used to investigate the nicotinic acetylcholine receptor from the ray *Torpedo*³¹ and to generate a three-dimensional reconstruction of frozen-hydrated tubes of receptors. These showed that the molecule is more elongated and angular than suggested by negative staining³². The reconstruction of the frozen-hydrated material showed that the molecule is a 6.5-nm diameter cylinder composed of five rod-like subunits, each ~14 nm long, arranged approximately perpendicular to the plane of the membrane, with more material located on the synaptic surface than at the cytoplasmic surface. The reconstructions also showed that the subunits are arranged with almost perfect pentagonal symmetry so that they are located in approximately equivalent positions around the central channel of the molecule. An unexpected feature is the general similarity between the different subunits that form the receptor³¹. Human bladder membrane¹⁰ and Omp C protein³³ were also well preserved in ice and similarly gave views of the portion of the protein subunits located in the membrane in addition to the external material seen by negative staining.

Several crystalline protein sheets have been examined in vitreous ice, but, at the resolutions of ~2 nm usually obtained, the differences between this material and negatively stained specimens were usually small^{8,34,35}. Similarly, crystalline sheets of ribosomes¹⁰ resembled, to a resolution of ~5 nm, material examined in glucose. But frozen-hydrated DNA crystalline sheets show preservation to ~0.35 nm and images where the data extend to 0.65 nm have been recorded indicating that the crystal DNA may have a novel conformation³⁶.

Frozen sections

Whole cells and tissues are generally too thick to be examined directly by electron microscopy and so it is necessary to cut sections through them. Freezing usually requires a cryoprotectant, such as sucrose or glycerol, to inhibit the formation of ice crystals produced by the slower cooling of a more bulky specimen. Material frozen in this way can also be used for complementary studies using fracturing and etching³⁷ or freeze substitution, although crystallization of ice during etching may occur. Although this technique has limitations, a range of objects have been examined. Several cryosectioned bacteria, for example, seem well preserved and these studies indicate that the mesosome often observed in plastic-embedded material is probably an artefact associated with fixation in osmium tetroxide³⁸. Cryosections of insect flight³⁹ and rabbit skeletal (M.S., A. McDowall and J. Dubochet, unpublished) muscle in

rigor (nucleotide-free solution) give optical diffraction patterns closely resembling the corresponding X-ray diffraction patterns of bulk specimens. Unfortunately, sectioning is much more difficult when material is frozen in vitreous ice than for specimens embedded in plastic, particularly if a cryoprotectant has not been used. Cutting artefacts are common and cryosections are often grossly distorted. Unlike plastic sections, the cutting distortions cannot be easily relaxed after sectioning and so the section usually undulates or is crushed into a fine network of rippling distortions or 'crevasses'^{40,41}. In addition, long-range order is lost as a result of these distortions and this makes signal-to-noise enhancement by Fourier-based image processing difficult. However, it might be possible to overcome the technical problems associated with freezing and sectioning, in which case the potential of this method to preserve cells in their native state would assume considerable importance.

Problems outstanding

Considerable new information has been generated for a range of frozen-hydrated biological specimens and this has been particularly rewarding with highly ordered objects, such as arrays in membranes, and in fragile objects that are poorly preserved using other methods. Even very delicate objects, such as lipid vesicles, do not seem to be greatly distorted^{20,28}. For viruses and many other objects, data from frozen-hydrated material compare very favourably with those from other sources and structure often seems better preserved than by negative staining or other methods. To this extent, the promise of vitreous ice to provide a less hostile environment seems to have been fulfilled. It also seems to cause fewer distortions on freezing than crystalline ice, particularly in more open samples¹⁰.

However, there are still several problems relating to the preservation of the specimen, the recording of images and the effects of radiation damage. Before freezing, surface tension in the liquid film can displace robust structures and crush delicate ones if the film is too thin^{10,28}. Also, evaporation from the thin aqueous film before it is vitrified can cause the salt concentration to rise substantially²⁴ which can produce difficulties with salt-sensitive specimens, although this problem might be reduced by working in a high-humidity environment. It is important to keep the specimen below 130 K because vitreous ice recrystallizes at higher temperatures¹². Freeze-drying artefacts can be mistaken for frozen-hydrated material and it is therefore prudent to record electron-diffraction patterns to show the presence of vitreous (amorphous) ice¹²; but this does not exclude material at the ice surface becoming freeze dried, nor ice condensing on a freeze-dried portion. A more reliable solution is to warm up the grid in the electron microscope (after photographing the specimen). This produces freeze-dried material and allows comparison to be made. The contrast in freeze-dried specimens is generally higher and fine details are often less distinct than in ice. Also the bubbling associated with radiation damage (see below) is not seen in most freeze-dried specimens.

The resolution obtained in images of frozen-hydrated material is often substantially lower than indicated by electron diffraction, which is probably a result of specimen movement during the recording of micrographs. This movement seems to derive from both mechanical and radiation-induced effects and is a major factor that limits the resolution of frozen-hydrated images to the same order as that obtained by negative staining and similar techniques. Mechanical movement can be reduced by the production of better cold stages, but radiation-induced specimen movement⁴² may represent a severe obstacle to high-resolution imaging.

Radiation sensitivity

Radiation damage can often become a limiting factor, depending on the resolution required in the image. If high resolution (better, say, than ~2 nm) is desired, the same conditions seem to apply as for specimens embedded in glucose or negative stain: the

electron dose must be kept below ~ 100 electrons per nm^2 to preserve secondary and tertiary structure. Large arrays of molecules must therefore be available to overcome the poor statistical definition of individual particles. The radiation sensitivity of samples in ice or other materials such as glucose appears approximately equal at this resolution, and decreases with decreasing temperature^{7,43}. At lower resolution, radiation damage first appears as bubbles around the biological particles in the vitreous ice. The onset of bubbling is apparently related to the amount of organic material present and generally occurs at a dose of 2,000–5,000 electrons per nm^2 . The precise mechanism is not well understood, despite much work, although inelastic scattering of electrons by the ice could contribute free radicals that subsequently attack biological material^{44–47}.

The intrinsic radiation sensitivity of frozen-hydrated biological material means that images have to be recorded with electron doses much lower ($10\text{--}100\times$) than those used on negatively stained specimens. This limits the resolution that can be obtained for different classes of object. The statistical definition of low-dose images is poor and, when compounded with the low intrinsic contrast of frozen-hydrated biological material, this produces images that generally have a poor signal-to-noise ratio. This has important implications for the sort of information that can be obtained. Broadly, three levels of structure are accessible by electron microscopy: the position of a particle; gross features of its shape, size and orientation; and its fine structure (the internal arrangement of secondary and tertiary structural elements). No preparative method seems likely to provide atomic information on single biological molecules, but the low signal from a frozen-hydrated object the size of most protein molecules makes it difficult even to detect the object at all; it is only with larger structures, such as ribosomes or viruses, that one could hope to get reliable indications of shape or orientation. Other methods, such as shadowing, that give stronger signals can often give more information about single molecules. Therefore, most work in ice has been directed towards regular macromolecular arrays, particularly crystalline sheets and particles with helical or icosahedral symmetry, because image processing can be used with these objects to improve the signal-to-noise ratio and allow detection of finer structural details. Even then, positional information is probably more reliable than that about size and shape, and so these methods tend to be most useful when examining the arrangement of subunits, or differences in structure arising from molecular rearrangements or the presence of labels. These problems also arise with negative staining, but the low signal-to-noise ratio of frozen-hydrated specimens greatly exacerbates the situation.

In principle, frozen-hydrated crystals can be used to examine molecular detail to near atomic resolution in the same way as material in glucose or negative stain, and because contrast at high resolution does not depend on density differences between object and solvent there is essentially no difference in contrast between samples prepared in different ways. However, there is compelling evidence from electron diffraction^{5,19,34,36} that crystalline order is extremely well preserved in many frozen-hydrated specimens and this, together with a reduction in radiation damage at low temperatures, may be a significant advantage for high-resolution work.

Interpretation of images

The problems of interpretation of micrographs from frozen-hydrated specimens are somewhat different from those encountered in other preparative techniques. In images of negatively stained material photographed at ~ 300 nm under focus (Fig. 5), spatial frequencies present in the object are represented with approximately equal fidelity to $\sim 2\text{-nm}$ resolution. But frozen-hydrated biological specimens have very little amplitude contrast or mass-thickness contrast because ice and protein have similar scattering cross-sections⁴⁸. Therefore, images are formed almost entirely by phase contrast, which is very low at focus

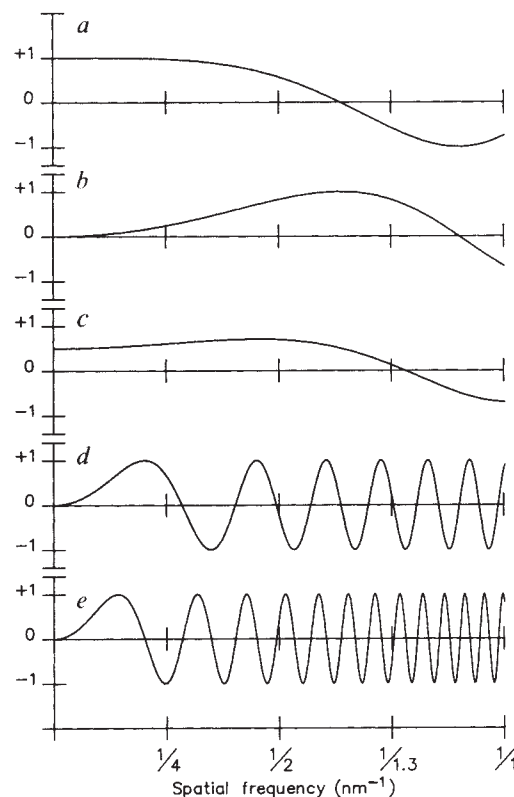


Fig. 5 Contrast-transfer functions for different imaging conditions in the electron microscope. Image formation in an electron microscope is conveniently thought of in terms of a contrast-transfer function that modulates spatial frequencies. This corresponds to multiplying the Fourier transform of the object by the transfer function and then synthesizing the image by inverse transformation. The transfer functions arise as a consequence of the mechanisms of image formation in the microscope which are discussed in detail elsewhere^{50–52}. Mass-thickness contrast derives from electrons scattered outside the objective aperture⁵², whereas amplitude and phase contrast arise from electrons transmitted by it^{50,51}. Amplitude contrast is strong at low spatial frequencies and a shows typical amplitude-contrast transfer function for an under-focus of ~ 300 nm. Phase contrast at this defocus (*b*) is weak at low spatial frequencies but then increases to a maximum at $\sim 1/(1.5\text{ nm})$. In negatively stained specimens (*c*), amplitude and phase contrast have comparable strength and complement one another so that spatial frequencies to $\sim 1/(2\text{ nm})$ are all transferred with approximately the same fidelity. But amplitude contrast and mass-thickness contrast are weak in frozen-hydrated samples and so images are formed mainly by phase contrast. Because of this, a greater defocus is used to bring the maximum of the phase-contrast transfer function to near $1/(5\text{ nm})$ – $1/(3\text{ nm})$ to maximize those spatial frequencies that are most important in defining the general position and outline of the object. Phase-contrast transfer functions for an underfocus of 3,000 and 6,000 nm respectively, which are typical of the values used for frozen-hydrated material, are shown in *d* and *e*. Although spatial frequencies in a range of $\sim 1/(10\text{ nm})$ to $1/(2\text{ nm})$ are transmitted with reasonable fidelity, low frequencies are transmitted poorly, whereas higher frequencies are substantially altered and, because the transfer function oscillates, are often included with the wrong sign. Unlike the light microscope, where the interference between scattered and unscattered (central) beams that generates the phase image is produced by a phase plate that retards the unscattered beam, the electron microscope produces this interference by defocusing the image. This produces the correct phase difference only for beams scattered at defined angles; beams at other angles will have their phase altered to a greater or lesser extent and so will not interfere with the central beam. Consequently, only the spatial frequencies that correspond to these angles will be represented with full weight. Therefore, it is not possible to record images of this material that represent all spatial frequencies in the object equally; instead images are obtained that only record part of the data from the image with high fidelity.

but increases with defocus. Micrographs of frozen-hydrated material are therefore usually recorded several micrometres under focus³⁵. Unfortunately, not all spatial frequencies originally present are then represented equally in the image (Fig. 5) which may alter details substantially, particularly fine features and boundaries of particles, and leads to difficulties of interpretation. Image processing can correct for many defocus-related errors, but there are difficulties when phase contrast is weak and the correction is correspondingly large. The Fourier transform of the image at these frequencies will contain substantial noise, and at low frequencies the amplitude contrast may be appreciable relative to phase³⁵. Furthermore, the high degree of inelastic scattering by ice may complicate image formation at low spatial frequencies. Therefore, under these conditions, simply multiplying by the inverse of the phase contrast transfer function may be ill-advised. In many instances, information on low spatial frequency components of the structure may have to be obtained from other sources, such as X-ray scattering or negative staining. Corrections for thick objects are even less straightforward because some of the assumptions used to model phase contrast break down. It is not clear how such images should be interpreted in detail.

Fortunately, the phase nature of the image is not usually a major problem with crystalline objects because they generally have little structural information at low spatial frequencies. Also, electron-diffraction amplitudes (unaffected by the phase-contrast transfer function) are often available. However, bounded objects, such as helical or icosahedral particles, may sometimes present difficulties because important features of their structure are often specified by low spatial frequency terms near the origin of their Fourier transform. In the case of helical objects, the equator specifies the overall radial density distribution in the particle and so determines the relative weight of features at different radii⁴⁹. It therefore influences substantially the boundaries chosen for both the whole particle and its constituent subunits. However, the equator is usually dominated by low-frequency terms that are substantially altered by the phase-contrast transfer function. Several methods have been used to address this problem. Corrected layer lines and an uncorrected equator can be used¹⁷, but this almost certainly underweights the equator. (This may not be completely undesirable as it sharpens edges, but it could produce difficulties in deciding connectivity or subunit boundaries.) Alternatively, equatorial data can be extrapolated using methods commonly used in low-angle X-ray diffraction. This was found to be satisfactory with microtubules²³, where most density is located at a single radius, but may be less reliable when density is more broadly distributed. A model equator, based on X-ray scattering and other data, was used for actin²⁴, but this sort of data is often not available and there can be problems with scaling.

Conclusions

The examination of frozen-hydrated biological material is not only feasible, but offers significant advantages for many specimens. Freezing samples in vitreous ice maintains them in an environment close to that in which they normally function and so minimizes artefacts from dehydration and staining while preserving structure to high resolution. This method also has the advantage of looking directly at the object, whereas other electron microscope methods infer boundaries from the exclusion of stain or evaporated metal. Therefore, artefacts associated with positive staining or failure of the stain to

delineate the particle surface closely are avoided. New data produced from frozen-hydrated specimens have already made a significant impact in structural cell biology, particularly in the study of arrays and membranes.

Many objects are reasonably well preserved to a resolution of 2–3 nm by conventional procedures such as negative staining, and so examination of this sort of material in ice is unlikely to yield much new structural information. Viruses, cell walls, cell surface and extracellular proteins, for example, are often subjected to harsh environments in which ionic strength and pH are poorly controlled and where dehydration is common. Therefore, it is not surprising that, for these particles, few differences are seen between negatively stained and frozen-hydrated material. There are, however, many other objects that are not so robust. For example, those from cell cytoplasm, where conditions are very strictly regulated, have always been thought likely to be altered by traditional preparative methods. Analysis of frozen-hydrated specimens provides a powerful means to assess these differences, which are often surprisingly small. Hence, one effect of cryo-electron microscopy is to increase confidence in negative staining techniques.

However, there are also many instances where the structure of biological material is clearly not well preserved using conventional techniques, even to a resolution much lower than 3 nm, and it is in these instances that examination in ice has been most rewarding and will probably continue to be so. This is particularly true for specimens that have some degree of ordering, so that structures can be more easily recognized and the signal-to-noise ratio enhanced by image processing. The high-level organization of chromatin is a topical example of this sort of problem. Examination in ice also offers an opportunity to exploit density differences in objects, both between gross structural features such as protein and DNA or lipid, and between the bulk of the object and small heavy-metal labels. Systems that undergo structural changes are particularly attractive objects, as illustrated by the work on calcium-induced changes in the structure of gap junctions. The ability to alter the composition of the medium, in terms of salt, ions such as calcium, ATP or even electrical potential, coupled with the ability to freeze quickly and so to preserve intermediates in conformational changes or assembly processes, may allow the integration of structural and biochemical data. The changes involving myosin and dynein in motile systems and the kinetics of assembly of tubulin, actin and surface layers are areas likely soon to yield exciting information.

It is also significant that in many instances electron diffraction has indicated that material in ice is preserved to very high resolution, often to better than 0.5 nm, although it is currently extremely difficult practically to obtain images to better than ~2 nm resolution. However, if it proves possible to overcome technical limitations related to specimen movement and radiation damage, the high degree of preservation in vitreous ice may form the basis of examination of biological material at atomic resolution in conditions approximating to those found *in vivo*.

We thank Jacques Dubochet, Jean Lepault, John Trinick and Nigel Unwin for supplying micrographs and information of material in press and to our colleagues, particularly Tom Ceska, Bob Glaeser, Richard Henderson, Hugh Huxley, Aaron Klug, Andrew Somlyo, Ken Taylor and John Trinick for comments and criticisms on the manuscript. G.V. holds an MRC research studentship.

1. Hoppe, W. *Naturwissenschaften* **55**, 65–74 (1968).

2. Parsons, D. F. *Science* **186**, 407–414 (1974).

3. Dubochet, J. *et al. J. phys. Chem.* **82**, 6727–6723 (1984).

4. Fernandez-Moran, H. *Expl Cell Res.* **3**, 282–350 (1952).

5. Taylor, K. A. & Glaeser, R. M. *Science* **186**, 1036–1037 (1974).

6. Taylor, K. A. & Glaeser, R. M. *J. ultrastruct. Res.* **55**, 448–456 (1976).

7. Glaeser, R. M. & Taylor, K. A. *J. Microsc.* **112**, 127–138 (1978).

8. Glaeser, R. M., Chiu, W. & Grano, D. *J. ultrastruct. Res.* **66**, 235–242 (1979).

9. Lepault, J., Booy, F. P. & Dubochet, J. *J. Microsc.* **129**, 89–102 (1983).

10. Milligan, R. A., Brisson, A. & Unwin, P. N. T. *Ultramicroscopy* **13**, 1–10 (1984).

11. Dubochet, J. & McDowell, A. D. *J. Microsc.* **124**, RP3–4 (1981).

12. Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A. & Homo, J.-C. *J. Microsc.* **128**, 219–237 (1982).

13. Bruggeller, P. & Mayer, E. *Nature* **288**, 569–571 (1980).

14. Jaffe, J. S. & Glaeser, R. M. *Ultramicroscopy* **13**, 373–378 (1984).

15. Chang, C.-F., Ohno, T. & Glaeser, R. M. *J. electr. microscop. Tech.* **2**, 59–65 (1985).

16. Adrian, M., Dubochet, J., Lepault, J. & McDowell, A. D. *Nature* **308**, 32–36 (1984).

17. Lepault, J. & Leonard, K. *J. molec. Biol.* **182**, 431–441 (1985).

18. Amos, L. A. & Klug, A. *J. molec. Biol.* **99**, 51–73 (1975).

19. Lepault, J. *J. Microsc.* **140**, 73–80 (1985).

20. Lepault, J., Pattus, F. & Martin, N. *Biochim. biophys. Acta* **820**, 315-318 (1985).
21. Booy, F. P., Ruigrok, R. W. H. & van Bruggen, E. F. J. *J. molec. Biol.* **184**, 667-676 (1985).
22. Earnshaw, W. C. & Casjens, S. R. *Cell* **21**, 319-331 (1980).
23. Mandelkow, E.-M. & Mandelkow, E. *J. molec. Biol.* **181**, 123-135 (1985).
24. Trinick, J., Cooper, J., Seymour, J. & Egelman, E. H. *J. Microsc.* **141**, 349-360 (1986).
25. Stewart, M. & Lepault, J. *J. Microsc.* **138**, 53-60 (1985).
26. Stewart, M. *J. molec. Biol.* **148**, 411-425 (1981).
27. Stewart, M. *Proc. R. Soc. B190*, 257-266 (1975).
28. Dubochet, J., Adrain, M., Lepault, J. & McDowell, A. D. *Trends biochem. Sci.* **10**, 143-146 (1985).
29. Unwin, P. N. T. & Ennis, P. D. *Nature* **307**, 609-611 (1984).
30. Unwin, P. N. T. & Ennis, P. D. *J. Cell Biol.* **97**, 1459-1466 (1983).
31. Brisson, A. & Unwin, P. N. T. *Nature* **315**, 474-477 (1985).
32. Brisson, A. & Unwin, P. N. T. *J. Cell Biol.* **99**, 1202-1211 (1984).
33. Chang, C.-F., Mizushima, S. & Glaeser, R. M. *Biophys. J.* **47**, 629-640 (1985).
34. Cohen, H. A., Jeng, T. W., Grant, R. A. & Chiu, W. *Ultramicroscopy* **13**, 19-26 (1984).
35. Lepault, J. & Pitt, T. *EMBO J.* **3**, 101-105 (1984).
36. Downing, K. H. *Ultramicroscopy* **13**, 35-46 (1984).
37. Heuser, J. E. *J. Cell Biol.* **84**, 560-583 (1980).
38. Dubochet, J., McDowell, A. D., Menge, B., Schmid, E. N. & Lickfeld, K. G. *J. Bact.* **155**, 381-390 (1983).
39. McDowell, A. W., Hofmann, W., Lepault, J., Adrain, M. & Dubochet, J. *J. molec. Biol.* **178**, 105-111 (1984).
40. Chang, J. J. *et al. J. Microsc.* **132**, 109-123 (1983).
41. McDowell, A. D. *et al. J. Microsc.* **131**, 1-9 (1983).
42. Henderson, R. & Glaeser, R. M. *Ultramicroscopy* **16**, 139-150 (1985).
43. Lamvik, M. K., Kopf, D. A. & Robertson, J. D. *Nature* **301**, 332-334 (1983).
44. Unwin, P. N. T. & Nuguruma, J. J. *J. appl. Phys.* **42**, 3640-3641 (1971).
45. Talmon, Y., Davis, H. T., Sriver, L. E. & Thomas, E. L. *J. Microsc.* **117**, 321-332 (1979).
46. Zeitler, E. (ed.) *Cryomicroscopy and Radiation Damage* (North Holland, Amsterdam, 1982).
47. Dubochet, J., Chang, J.-J., Freeman, R., Lepault, J. & McDowell, A. D. *Ultramicroscopy* **10**, 55-62 (1982).
48. Eusemann, R., Rose, H. & Dubochet, J. *J. Microsc.* **128**, 239-249 (1982).
49. Klug, A., Crick, F. H. & Wyckoff, W. W. *Acta cryst.* **11**, 199-213 (1958).
50. Erickson, H. P. & Klug, A. *Phil. Trans. R. Soc. B261*, 105-118 (1971).
51. Glaeser, R. M. *A. Rev. Phys. Chem.* **36**, 243-275 (1985).
52. Zeitler, E. & Barr, G. F. *Expl Cell Res.* **12**, 44-65 (1957).

ARTICLES

Occultation detection of a neptunian ring-like arc

W. B. Hubbard*, A. Brahic†, B. Sicardy†, L.-R. Elicer‡, F. Roques† & F. Vilas*§

* Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

† Université Paris VII, Observatoire de Paris, 92190 Meudon, France

‡ Cerro Tololo Inter-American Observatory, Casilla 603, La Serena, Chile

The apparent closest approach of the star SAO186001 to Neptune was observed photoelectrically on 22 July 1984 at Cerro Tololo Inter-American Observatory. A 32% signal drop lasting about 1.2 s was probably caused by a partially transparent arc of material at a distance of 67,000 km from Neptune. Neptune's arc(s) do not vary smoothly with azimuth, unlike the rings of other jovian planets.

IN this article we show for the first time that Neptune, like Saturn, Uranus, and Jupiter, possesses a ring-like structure. Although it has been argued that Neptune differs from the ringed planets because of its irregular satellite system, we now find that it lies closer to the norm. But all three previously detected ring systems are markedly dissimilar and Neptune is no exception.

The detection of a planetary ring is significant because it relays information about possible accretion, collision, and orbital resonance processes nearby, and motion of the ring under the influence of the planet's gravity acts as a probe of the planet's internal structure.

It is more difficult to use stellar occultations to search for neptunian rings than uranian rings for two basic reasons: Neptune is more distant than Uranus and thus its cross-section is about a factor of two smaller, and any equatorial ring is seen at a rather oblique angle. With the sensitivity of the technique used, material around Neptune can be detected to an optical depth limit of the order of 1%. Until now, rings observed and confirmed by occultation techniques have been composed of occulting material forming an azimuthally continuous ring around the planet. The stellar occultation technique has been effective in detecting previously unknown rings and low-orbit satellites, but it suffers from spurious signal interruptions. Many observed signal drops which have nothing to do with celestial occulting material can be explained by poor guiding, clouds, electronic problems, and so on, but sometimes an observation has no such obvious explanation. Although suspicious events have been reported around Uranus as well as around Neptune, only events confirmed by independent observations or showing identical structure before and after the centre of the occultation event can be considered reliable¹. The peculiar character of the

Table 1 Summary of recent occultation searches

Search no.	Date	Location(s)	Ref.
1	10 May 1981	Catalina Station, Arizona, San Pedro Martir, Baja California Norte	3
2	10 May 1981	Mauna Kea, Hawaii	6
3	10 May 1981	Mt Stromlo, Australia	6
4	15 June 1983	Chung-Li, Taiwan	4
5	15 June 1983	Mauna Kea, Hawaii	5
6	15 June 1983	Ayers Rock, Australia	4
7	15 June 1983	Hobart, Tasmania, Australia	4
8	24 May 1981	Cerro Tololo, Chile	6
9	24 May 1981	Catalina Station and Mt Lemmon, Arizona	7

neptunian ring system, which we discuss below, particularly demands a cooperative observational program for its investigation.

Figure 1 and Table 1 summarize published occultation searches from a systematic campaign carried out over the past 4 years to detect neptunian rings by stellar occultation. In Fig. 1 the line with tick marks shows the geometry of the appulse discussed here². Straight lines show the apparent path of a star behind the planet, while the dashed curve shows the approximate locus of points in an equatorial orbit at a distance of 75,000 km from Neptune's centre (about $3R_N$, where R_N is Neptune's radius). Information about each previous search³⁻⁷ is given in Table 1. Searches 4-7 have recently been augmented by more published data from the same occultation which likewise gave negative results⁸. Studies of a favourable occultation on 20 August 1985 are in progress^{9,27}. In only one previous case (search

§ Present address: NASA/Johnson Spacecraft Center, Houston, Texas 77058, USA.

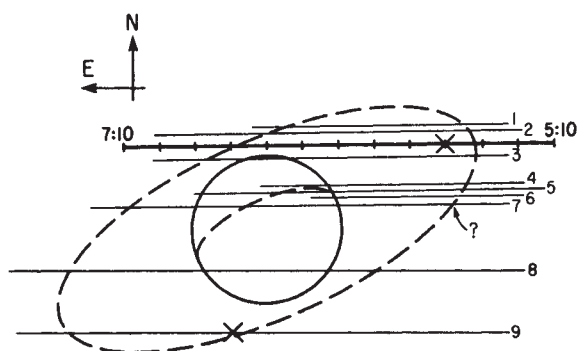


Fig. 1 Searches for neptunian rings identified in Table 1. The heavy line with tick marks every 10 min shows the apparent path of SAO 186001 behind Neptune from 5:10 to 7:10 UTC on 22 July 1984 as seen from Chile; upper $\times$ the detection discussed in this paper; lower $\times$, an earlier coincidence detection of an occulting object⁷. Dashed lines represent the equator and an equatorial orbit of radius 75,000 km.

9) was an occultation event confirmed by simultaneous observation from another telescope. Below we provide some additional information about the event reported from search 9, which suggests that this event may actually have been caused by a ring-like arc. Except for some isolated suspicious events (T. Gehrels and P. D. Nicholson, personal communications), results were otherwise uniformly negative, showing that Neptune possesses neither a broad continuous Saturn-style ring system nor narrow closed rings like Saturn's F-ring or those of Uranus. However, an optically thin ring like Jupiter's would remain undetected by Earth-based occultation techniques. The feature sought was a continuous ring of occulting material which would necessarily be visible twice on each chord (unless hidden by the planet).

Our conclusions are based on simultaneous observations made at two observatories on 22 July 1984. The observations presented here comprise one of the two data sets which are equally essential for revealing the nature and geometry of an object detected on that date. The combined observations demonstrate the existence of an arc of occulting material¹⁰⁻¹⁴. It is not clear whether the material comprises a complete ring. Any model for the material must explain why it has generally eluded detection in previous occultation searches, but is still obvious in our data. Either the system is discontinuous, or its optical depth varies greatly with azimuth. The geometry of the feature may even be completely novel.

Although we do not yet know a great deal about the characteristics of the arc system, a number of interesting theoretical problems are already posed by this detection, such as: influence of a massive, nearby, retrograde satellite, possibility of a broken and/or highly variable ring, and clumping and accretion of material near the Roche limit.

Experimental method

W.B.H. planned and scheduled an experiment to observe this appulse at Cerro Tololo Inter-American Observatory (CTIO). The CTIO experiment was on a single telescope, using the observatory's three-channel high-speed photometer. F.V. and L.-R.E. made the actual observations; for experimental details see Fig. 2 legend. Corresponding experiments were planned and arranged at European Southern Observatory (ESO) by A.B. and by B.S., although data from those observations are not presented here¹⁵. The telescope used at CTIO is about 102 km south and 7 km west of ESO.

Before the observation at CTIO, the observers experimented with filters in channels 1 and 3. The intention was to use Johnson V and B filters in these channels respectively to attempt to measure refractive dispersion in Neptune's atmosphere¹⁶. Although calibration observations of Neptune and the star sep-

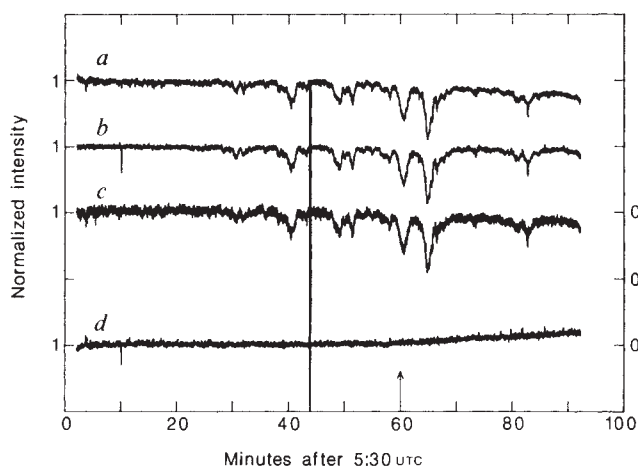


Fig. 2 Normalized light curves for: *a*, channel 1; *b*, channel 2; *c*, channel 3; *d*, normalized ratio of channel 2 counts to channel 1 counts. Full-scale drop at about 6:13 is a sky-background measurement. Arrow indicates the approximate predicted time for second crossing of the equatorial plane at a distance of 67,000 km. 0.9-m telescope; focal plane aperture = 0.77 mm (12.7 arc s). Channel 1: unknown bandpass, S-20 detector. Channel 2: 0.87 μm , $\delta\lambda = 0.08 \mu\text{m}$, RCA 31034A (III-V) detector. Channel 3: Johnson B-band, 0.44 μm , $\delta\lambda = 0.1 \mu\text{m}$, S-20 detector. All channels: integration rate = 100 Hz, dead time constant $\approx 4 \times 10^{-8}$ s. Weather: initially photometric, intermittent cirrus clouds after 5:55 UTC. Latitude: $-30^{\circ}10'07.5''$; longitude: $70^{\circ}48'22.4''$; altitude: 2,210 m. Observers: F.V., L.-R.E.

arately were made with the V filter in channel 1, this filter was apparently accidentally changed before the appulse observations. We are thus unsure what effective bandpass was used in channel 1.

Results of calibration observations are given in Table 3. The combined signal observed during the appulse was the sum of the separate star and Neptune calibration signals, except for channel 1, as explained.

Although we did not plan to set up coincidence experiments at separate observatories in advance, the geometry was ideal for demonstrating the reality of the detection which we report below.

Presentation of data

Figure 2 shows 1-s averages of the total intensity in the three channels as a function of time. In each case, the light curve is the combined star plus Neptune signal, normalized to unity at full intensity. The CTIO observations began at 5:32. The effect of clouds on the signals is noticeable after 5:55, but can be effectively removed by dividing channel 2 by channel 1 (lowest trace in Fig. 2). The gradual increase in the ratio after 6:30 is caused by differential extinction as the star and planet set.

The interruption of the signal at about 5:40 is striking and obviously has no counterpart anywhere else in the light curves. Because the channel 3 data provide continuous monitoring of an uninterrupted signal from Neptune which was at all times combined in the beam with the star, we can rule out terrestrial causes of the interruption such as clouds, aircraft, birds, electrical disturbances, or bad guiding. Triton and Nereid were not in the vicinity of Neptune. Thus, an unknown celestial object was responsible for the occultation. Data from a single station alone, such as CTIO or ESO, are not sufficient to rule out the possibility that the object was an asteroid or a single, small (~ 10 km) solid object in orbit around Neptune. But the combined data from the two stations, which are equally necessary for this purpose, show that the object could only have been a ring-like arc of partially transparent material.

There is no detectable occultation at any time other than 5:40. If the object which caused the occultation at 5:40 was an

Table 2 Calibration parameters for star and Neptune

	Channel 1	Channel 2	Channel 3
Star	59	306	3
Neptune	14	36	41
Combined	195		

All numbers are effective photon count rates per 10 ms time interval.

equatorial ring, this ring would have been crossed again at about 6:30.

The occultation event in channel 2 near 5:40 is shown at full time resolution in Fig. 3. The position of the star in the sky plane relative to Neptune would have a velocity of 20.1 km/s eastward and 0.22 km/s northward. According to the calibration data, a stellar drop equal to that observed in channel 2 would lead to a drop no greater than about 2.5 percent in channel 3. Thus the essential continuity of the signal in channel 3 confirms the reality of the stellar occultation, showing that the light from Neptune remained uninterrupted. A small decline ($\sim 5\%$) is marginally discernible in channel 1. Assuming that the fractional decline in the starlight in channels 1 and 2 was the same, we calculate that the 'correct' count rates from the star and Neptune in channel 1 were about 37 and 158 per 10 ms respectively (see Table 3). This would correspond to an effective wavelength between Johnson V and B and about a factor of 2 or 3 wider bandpass.

The published description of the occultation events observed with the 0.5-m and 1-m telescopes at ESO¹³ is remarkably similar to the event shown in Figure 3. From the similarity of the three occultation events over a 100-km baseline, it is clear that we have observed a ring-like feature rather than a single absurdly elongated satellite. In fact the events are qualitatively similar to profiles observed for the narrow and dense uranian rings^{17,18}. But despite the local similarity of the feature to a uranian ring, we emphasize that the lack of continuity around the planet shows that the occulting object is not a conventional ring.

The α -ring of Uranus is $\sim 50\%$ transparent¹⁷. Recent analysis^{18,19} has reached similar conclusions for the α -ring and several other uranian rings. Thus as a first approximation we model the neptunian occulting object as a homogeneous bar of length ≥ 100 km, projected width on the sky plane W and effective transparency f_0 . The projected position angle of the bar is computed from the timings at ESO¹³ and CTIO; the velocity component of the star normal to the projected bar is 18.5 km s^{-1} . The full wave-optical occultation profile is then computed and convolved over the projected stellar disk¹⁷. Assuming that SAO186001 is a K5 supergiant with $V=9$, we compute a projected diameter of 10 km at Neptune. After convolution over the stellar disk, the profile is nearly independent of diffraction effects.

Least-squares fitting of the above model to the data gives $W=15.0$ km, $f_0=0.68$, and central time $t_0=5:40:08.82$. The uncertainty in the central time is to some extent model-dependent, but does not exceed about 0.03 s. The duration of the event will depend on the larger of the bar width and the stellar diameter. If the bar is opaque, it must be smaller than the stellar diameter in order to fit the maximum signal drop. Thus for the calculated projected stellar diameter (10 km), an opaque bar cannot fit the data with a 32% dip. Figure 4 shows comparisons of models with normalized high-resolution data from channel 2. The best-fit model ($W=15$ km, $f_0=0.68$) is curve *b*. Curve *c* corresponds to the least-squares fit for an opaque bar. The best fit has been obtained for $W=3.7$ km. Curve *a* corresponds to a more transparent bar ($f_0=0.8$), for which the best-fit $W=25$ km.

Obviously type *a* models (transparency > 0.7) are unsatisfactory. Type *c* models (opaque bar) can give the correct central depth, but it is difficult to fit the wings to the data. This is why,

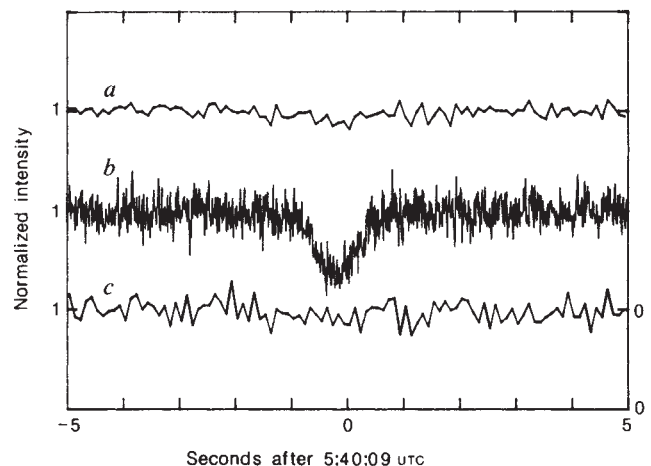


Fig. 3 Expanded view of occultation event. *a*, channel 1, averaged at 0.1 s intervals and full signal normalized to unity; *b*, channel 2 a full time resolution of 0.01 s, with the calibrated Neptune signal subtracted and the remaining unocculted stellar signal normalized to unity; *c*, channel 3, averaged at 0.1 s intervals and full signal normalized to unity.

as a first approximation, we prefer type *b* models (transparency close to 0.7). Taking into account the inclination of the ring plane relative to the line of sight, a normal optical depth of the arc of 0.14 can be deduced from a transparency of the order of 0.7. Note that, in this case, the calculated stellar diameter of 10 km is smaller than the bar and the profile should exhibit a flat bottom. The data do not show such a feature. If the bar is made narrower to eliminate the flat bottom, the entire feature becomes too narrow. However, if the stellar diameter is included as an adjustable parameter, a better fit to both the wings and the bottom of the dip is given with a stellar diameter of 21.8 km ($W=14.8$ km; $f_0=0.62$; $t_0=5:40:08.82$). Furthermore, the transparency of the arc may vary across its width, gradually increasing towards the edge (in contrast, a sharply confined ring is expected to have a profile with sharp edges, with maximum density at the edges). This matter merits further study.

Astrometric solution

Because Neptune did not occult the star, we have no direct determination of the position of the occultation points with respect to Neptune. Our calculation of these points is based upon results from three astrometric plates of Neptune and SAO186001 on 8 June 1984, which were taken and measured by A. Klemola (personal communication relayed by R. L. Millis), when the objects were a little more than 1° apart. We used these results to compute corrections to the ephemeris of Neptune relative to the star, with the resulting path for CTIO shown in Figure 1. Typically such corrections have an uncertainty of about ± 0.15 arc s in both right ascension and declination. At Neptune's distance, this corresponds to an uncertainty projected on the sky plane of about $\pm 3,000$ km in both coordinates. We have increased the error bar in right ascension to $\pm 4,000$ km to allow for an estimated larger uncertainty in the correction to the ephemeris in this coordinate.

From the ESO¹³ and CTIO timings, we compute the separation of the ESO and CTIO occultation points on the arc to be 100 km at a position angle of 22.4° . Figure 5 shows that, considering the probable uncertainties in the location of the occultation points with respect to Neptune, and in the position of Neptune's pole, this result is consistent with a circular arc in Neptune's equatorial plane, concentric with the centre of the planet. If we plot the locus of all points on concentric equatorial circles where the local tangent line has a projected position angle in the sky plane of 22.4° , the locus forms a straight line which should pass through the error box of the occultation point on the sky plane,

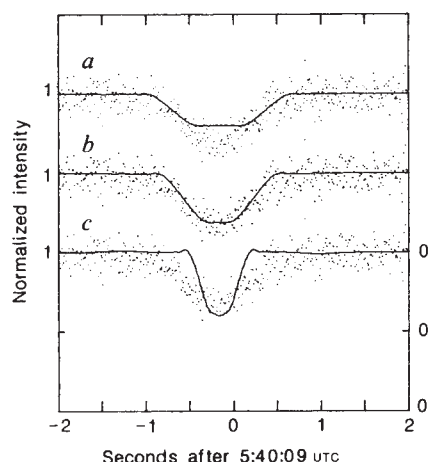


Fig. 4 Theoretical profiles (curves) compared with normalized signal from the star in channel 2 (points). See text for details of models.

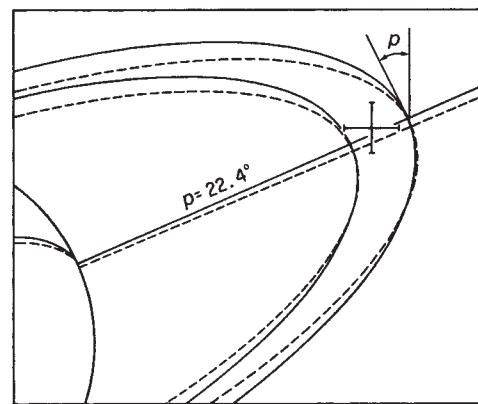


Fig. 5 Sky-projection of the limb of Neptune and equatorial points at distances of 63,000 and 71,000 km from Neptune's centre (ellipses). Solid curves are calculated using Harris' Neptune pole position with nominal Triton mass, while dashed curves correspond to zero Triton mass²⁶. Error bars designate the computed position of the occultation events with respect to the centre of Neptune. p , position angle.

as is evidently the case in Figure 5. The arc is 63,000–71,000 km from the planet's centre (estimated range of uncertainty). As it was apparently observed at the ansa, the projected width of 15 km is essentially equal to its true radial width.

It is an important constraint on any model that most occultation observations (including ours on the other side on 22 July 1984) show no detectable material. For this reason, it is important that previous observations be carefully searched for possible low-level events, and that future observations be coordinated for coincidence detections. Already some previously unreported possible detections have been noted (ref. 12 and T. Gehrels and P. D. Nicholson, personal communications) which are roughly consistent with the $3R_N$ value reported here. We caution that some unconfirmed detections may not be genuine. In all cases, however, the occulting object does not fit any readily discernible pattern (for example, there is no evident correlation with Triton's position); its width and transparency apparently can vary by an order of magnitude, and a continuous ring with properties varying smoothly with azimuth appears to be ruled out. The 1981 detection from near Tucson⁷ has been reinterpreted²⁰ and may be either a very dense arc or a satellite which is intimately related to the arc. The principal evidence that it is an arc consists of the calculated position angles of the immersion and emersion occultation points from the two stations, which agree with the predicted position angle of an equatorial ring segment. Non-detection of the 1981 object at Flagstaff, Arizona, constrains its length to be less than a few hundred km if it is an arc.

Theoretical implications

With such a limited data base, it is much too early to present a complete model of Neptune arc(s). Nevertheless, we can consider a list of possible geometries for the material detected by occultation. As we have shown, the alignment and elongation of the arc is consistent with dispersed material in an equatorial orbit around Neptune, and unless it is coincidental, argues that the feature is dynamically associated with Neptune.

Note that the arc is in Neptune's equatorial plane, inside the classical Roche limit (for a density of 1 g cm^{-3} for the ring material) and near the upper accretion limit (inside of which the gravitational attraction for two equal-size spherical particles is smaller than Neptune's tidal force).

Probably what we have seen is a part of a more complex ensemble; an interpretation of the data must take account of the following points:

(1) An unconfined ring will lose energy because of interparticle collisions while conserving angular momentum, leading to slow

spreading. But even for the feature which we observe, the spreading time is short compared with the age of the solar system (depending on the particle size). Such an unconfined ring must be circular and continuous.

(2) The observations can be explained by a broken ring. At our present level of understanding, each occultation observation is thus a gamble, which may or may not lead to a detection (the success rate appears to be only about 10%). Although azimuthally discontinuous, kinky ringlets in the Encke division have been observed²¹, there have been no theoretical predictions of such arc-like structures; they may bear a relation to satellite accretion processes and clumping of material near the Roche limit. The arc structure may represent an intermediate stage between a traditional ring and a fully-formed satellite, possibly resulting from dynamical effects of Neptune's figure, Triton, and undetected material, making the arc part of a complex system.

(3) If the arc is part of an unbroken ring, the optical depth of the ring must vary by orders of magnitude over small (~ 1000 km) azimuthal scales, judging from a recent reinterpretation of the 1981 event²⁰. Note that a 200-m-wide opaque segment would be as hard to detect as a 100-km-wide low-optical depth segment. This differs fundamentally from the situation observed for Saturn and Uranus^{21,22}, where one finds rings with smooth variations in width and optical depth, and which precess regularly about the planet.

(4) In principle, a satellite which recently entered within the Roche limit could now be in the process of destruction by tidal forces, leading to incomplete ring structures. However, this hypothesis requires that we are living in a special epoch. Although the ring/arc structure could be in a long-term steady state but rapidly varying with time at any instant (similar to a planetary atmosphere), this is not supported by any other observation or theory at present.

(5) We cannot exclude the possibility that the arc material is confined near the vicinity of the L4 or L5 stable points on the orbit of an unknown satellite. But analogous objects elsewhere in the solar system are accreted bodies (Trojan asteroids and Saturn's co-orbitals of Dione and Tethys). Lissauer²³ has a model of an azimuthally-confined arc involving a Lagrangian point of a 150-km satellite plus another 150-km shepherd satellite.

(6) Material on horseshoe orbits²⁴ would have precisely the opposite geometry from that required here: a generally filled ring, with rare gaps. Such a model is more appropriate to co-orbital satellites than to rings (e.g. Janus and Epimetheus).

(7) Neptunian arc(s), unlike the ring systems of the other jovian planets, will be subjected to the gravitational influence of a nearby, massive, retrograde, inclined satellite, Triton. This represents an interesting dynamical problem which merits further exploration. The effect will be to thicken the ring vertically, change the ring plane and change the ring precession rate. Horizontal structure may be affected by resonance, but no low-order resonances appear to be operating.

(8) Froeschlé and Scholl²⁵ have shown that a resonance with a yet-unknown satellite can produce eccentric and inclined arcs. The stability of such arcs remains to be studied.

Conclusions

It is remarkable that systematic occultation campaigns (particularly in 1981 and 1983) yielded no ring detections, while chance observations in 1984 have produced a definitive measurement. Neptune's ring-like system is elusive and still poorly understood. Clearly still more intensive, coordinated occultation observations must be carried out. But the problem is of more than scientific urgency. The Voyager 2 spacecraft must pass through Neptune's equatorial plane in 1989 at a point very close to $3R_N$ in order to execute its planned encounter sequence with Neptune and then Triton. Every available occultation opportunity, possibly involving uncatalogued infrared stars and fairly

distant appulses which include the $3R_N$ region, should be considered for observation. Coordinated coincidence observations over scales of 1,000 km as well as 100 km will be important. As is evident from the circumstances of this discovery¹⁵, every effort must be made to ensure that the observations are carried out in a carefully coordinated, collaborative manner, with rapid pooling and dissemination of the results. Global support from established observatories will be important.

If the material in Neptune's arcs is as sparse as presently indicated, direct imaging of the material from the earth will be difficult. Imaging of the much denser Uranus rings is already marginal. But in principle such measurements would be desirable to get constraints on the albedo and composition of the particles, as well as their mass distribution.

This research was supported by NASA Grant NSG-7045 (Planetary Astronomy Program), and by ESO, FNRS, INAG, CNRS, University of Paris and Observatory of Paris. F.V. was visiting astronomer, Cerro Tololo Inter-American Observatory, supported by the NSF under contract No. AST 78-27879. We thank P. Bouchet, A. Gomez, F. Gutierrez, M. Mouchet, O. Pizarro, and R. Vega for observations and technical assistance, and S. Dermott, R. Millis, and E. Spiegel for useful discussions. We also thank all colleagues who have stimulated and helped us with their comments.

Received 31 December 1985; accepted 15 January 1986.

1. Sicardy, B. *et al.* *Nature* (submitted).
2. Mink, D. J., Klemola, A. R. & Elliot, J. L. *Astr. J.* **86**, 135-137 (1981).
3. Hubbard, W. B. *et al.* *Bull. Am. astr. Soc.* **13**, 728 (1981).
4. Hubbard, W. B. *et al.* *Astr. J.* **90**, 655-667 (1985).
5. Sicardy, B. *et al.* *Bull. Am. astr. Soc.* **15**, 816 (1983).
6. Elliot, J. L. *et al.* *Nature* **294**, 526-529 (1981).
7. Reitsema, H. J., Hubbard, W. B., Lebofsky, L. A. & Tholen, D. J. *Science* **215**, 289-291 (1982).
8. Elliot, J. L. *et al.* *Astr. J.* **90**, 2615-2623 (1985).
9. Vilas, F. & Hubbard, W. B. *Bull. Am. astr. Soc.* **17**, 923 (1985).
10. *IAU Circ.* No. 3962 (1984).
11. *IAU Circ.* No. 3968 (1984).
12. *IAU Circ.* No. 4022 (1984).
13. Roques, F. *et al.* *Bull. Am. astr. Soc.* **16**, 1027-1028 (1985).
14. Hubbard, W. B. *et al.* *Lunar planet. Sci.* **16**, 368-369 (1985).

15. Maddox, J. *Nature* **318**, 505 (1985).
16. Elliot, J. L., Wasserman, L. H., Ververka, J. & Sagan, C. *Astrophys. J.* **190**, 719-729 (1974).
17. Hubbard, W. B. & Zellner, B. H. *Astr. J.* **85**, 1663-1669 (1980).
18. Elliot, J. L., French, R. G., Meech, K. J. & Elias, J. H. *Astr. J.* **89**, 1587-1603 (1984).
19. Nicholson, P. D., Matthews, K. & Goldreich, P. *Astr. J.* **87**, 433-447 (1982).
20. Hubbard, W. B. *Science* (in the press).
21. Cuzzi, J. N. *et al.* *Planetary Rings* (eds Greenberg, R. & Brahic, A.) (University of Arizona Press, 1984).
22. Elliot, J. L. & Nicholson P. D. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) (University of Arizona Press, 1984).
23. Lissauer, J. J. *Nature* **318**, 544-545 (1985).
24. Dermott, S. F. & Murray, C. D. *Icarus* **48**, 1-11 (1981).
25. Froeschlé, C. & Scholl, H. *Astron. Astrophys.* (submitted).
26. Harris, A. W. *NASA Conf. Publ.* **2330** 357-373 (1984).
27. Sicardy, B. *et al.* *Bull. Am. astr. Soc.* **17**, 923-924 (1985).

Cloning of cDNA encoding the murine IgG1 induction factor by a novel strategy using SP6 promoter

Yoshihiko Noma, Paschalis Sideras*, Takayuki Naito, Susanne Bergstedt-Lindquist*, Chihiro Azuma, Eva Severinsson*, Toshizumi Tanabe, Tatsuo Kinashi, Fumihiko Matsuda, Yoshio Yaoita & Tasuku Honjo

Department of Medical Chemistry, Faculty of Medicine, Kyoto University, Kyoto 606, Japan and * Department of Immunobiology, Karolinska Institute, Wallenberg Laboratory, Lilla Frescati, S-104 05 Stockholm 50, Sweden

Complementary DNA encoding the IgG1 induction factor, the first lymphokine directed to B lymphocytes, from a murine T-cell line has been cloned using a new strategy. The putative primary amino-acid sequence was deduced from the nucleotide sequence determined. The lymphokine synthesized by the direction of this cloned cDNA has many other functions, such as production of B-cell growth factor-1 and induction of Ia on B cells.

IMMUNOGLOBULINS are composed of light and heavy (H) chains, each consisting of variable (V) and constant (C) regions. V regions are responsible for antigen binding and associate with different C regions that exert various effector functions such as complement fixation. The V-region repertoire of the heavy chain is shared by the immunoglobulin classes and subclasses (IgM, IgD, IgG3, IgG1, IgG2b, IgG2a, IgE and IgA) which are determined by one of eight C_H genes organized in the following order in the mouse genome: C_{μ} , C_{δ} , $C_{\gamma 3}$, $C_{\gamma 1}$, $C_{\gamma 2b}$, $C_{\gamma 2a}$, C_{ϵ} and

C_{α} . A single B lymphocyte can associate a given V_H region with different C regions, the first with the C_{μ} region and subsequently with the C_{γ} , C_{ϵ} or C_{α} region; this phenomenon is termed class switching. We and others have shown that somatic recombination between switch (S) regions located 5' to each C_H gene except for the C_{δ} gene (S-S recombination) is responsible for the class-switch phenomenon¹⁻⁷. The S-S recombination results in deletion of the intervening DNA segments between two S regions^{1,6}.

The relative abundance of different immunoglobulin classes is dependent on the nature of the antigen. Thymus-independent activation of B cells *in vivo* or *in vitro* gives rise to stimulation of IgG3 antibodies⁸⁻¹⁰, whereas thymus-dependent stimuli generally give rise to IgG1 and IgG2 antibodies^{8,9}. Selective T-helper cells for certain immunoglobulins have been described, using conventionally immunized T cells or T-cell lines¹¹. Furthermore, soluble products secreted by some T-cell lines increased IgG1 production when these supernatants were added together with lipopolysaccharide (LPS)¹²⁻¹⁴. One putative lymphokine in these supernatants was B-cell IgG differentiation factor (BCDF γ)¹² or IgG1 induction factor¹⁴. It was not clear whether the IgG1 induction factor induces selective activation of B cells which were already committed to IgG1 production, or whether it directs the S_{μ} - S_{γ} recombination to increase IgG1-producing B cells. To solve these problems we set out to isolate the complementary DNA encoding the IgG1 induction factor and to obtain the pure material in a large quantity. Characterization of this lymphokine may also facilitate understanding of the roles of many other putative lymphokines involved in B-lymphocyte maturation¹⁵.

A major problem for cloning cDNA from a rare source such as cultured T-cell lines is the small concentration of messenger RNA. To circumvent this problem an expression vector system containing the simian virus 40 (SV40) promoter was developed and successfully applied^{16,17}. However, this vector system still had some other problems to be solved. First, the expression vector was introduced to COS cells by DNA transfection, the efficiency of which varied depending on the DNA. Second, efficiency of product formation by COS cells was not always satisfactory. In addition, secreted products were diluted by culture media and contaminated by serum proteins. Third, products which were not secreted from COS cells were difficult to detect by their activities because there are many other inhibitory proteins in COS cell cytoplasm. Fourth, many COS cells are contaminated by mycoplasma, which often disturbs the biological assay. We have introduced a novel strategy that uses the high efficiency of *in vitro* transcription of the SP6 promoter^{18,19} and the high efficiency of translation in *Xenopus* oocytes. Using this method we report here cloning and characterization of cDNA encoding the murine IgG1 induction factor. This lymphokine was shown to express multiple biological functions such as B-cell growth factor 1 and Ia induction factor.

IgG1 induction factor mRNA

Mouse spleen cells given LPS *in vitro* become activated to secrete mainly IgM, IgG3 and IgG2b. The other isotypes, IgG1, IgG2a, IgA and IgE, are present only in small amounts or in quantities that are too low to be detected. Addition of the IgG1 induction factor to LPS-containing cultures resulted in a dose-dependent increase in the IgG1 isotypes with a concomitant decrease in the levels of IgG3 and IgG2b (refs 14, 20). Sideras *et al.*²⁰ recently described the development of new T-cell lines secreting the IgG1 induction factor. One of these lines, 2.19, secreted large amounts of the IgG1 induction factor when stimulated with concanavalin A (Con A), serving as a suitable source of mRNA specific for IgG1 induction factor.

We tested for the presence of mRNA encoding the IgG1 induction factor by *in vitro* translation in *Xenopus* oocytes. Poly(A)⁺ RNA of the 2.19 T-cell line, activated by Con A for 6 h, was injected into *Xenopus* oocytes and incubated for 36 h at 20 °C in Barth's medium²¹. The supernatants of the oocyte incubation medium were added to mouse spleen cells together with LPS. The incubation medium of oocytes injected with poly(A)⁺ RNA (1 μ g) of the 2.19 T-cell line elevated the IgG1 response about sixfold compared with cultures with LPS alone (data not shown). The stimulation of the IgG1 induction by oocyte translation products of 2.19 T-cell mRNA was as strong as that of the culture supernatant of the 2.19 T cells. As controls, incubation media of oocytes injected with phosphate buffer or myeloma mRNA did not stimulate IgG1 synthesis.

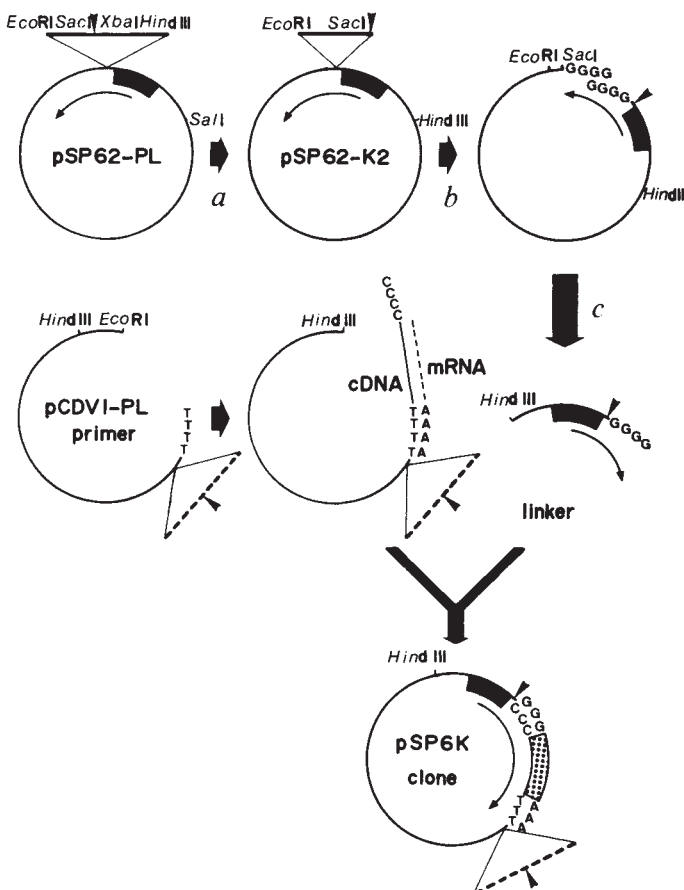


Fig. 1 Strategy for the construction of the SP6 expression vector library. Closed and dotted columns show SP6 promoter and double-stranded cDNA, respectively. Arrowheads show *Bam*HI sites. Arrows indicate the direction of transcription by SP6 polymerase.

Methods. Plasmid pSP62-PL (NEN), which contained a polylinker consisting of *Eco*RI, *Sac*I, *Sma*I, *Bam*HI, *Xba*I, *Sal*I, *Pst*I and *Hind*III sites (bold bar), was digested with *Xba*I and *Hind*III, treated with the Klenow fragment of DNA polymerase I and ligated to be circularized. The plasmid was digested with *Sal*I and treated with Klenow fragment. After ligation with a commercial *Hind*III linker (8 bp), the recombinant plasmid was digested with *Hind*III and ligated again. Plasmid pSP62-K2, which deleted the *Sal*I/*Pst*I/*Hind*III sites from the original polylinker and replaced the original *Sal*I site with a *Hind*III site, was thus constructed (a), digested with *Sac*I and poly(dG) tails of 14 bases (on average) per end were added to the *Sac*I termini (b). The DNA was then digested with *Hind*III and the dG-tailed linker DNA (~500 bp) containing the SP6 promoter was isolated by electrophoresis in a 1% agarose gel and recovered with DEAE-cellulose paper (c). A polylinker DNA fragment consisting of *Eco*RI, *Sac*I, *Sma*I, *Ava*I, *Bam*HI, *Xba*I, *Hinc*II, *Acc*I, *Sal*I, *Pst*I and *Hind*III sites (dashed bar) was isolated from pSP65 (Promega Biotec) and inserted into the *Bam*HI site of pCDVI. The plasmid formed was named pCDVI-PL. As the 5' protruding ends of the polylinker and *Bam*HI site of pCDVI were filled using Klenow fragment, there is neither *Eco*RI nor *Hind*III site in the polylinker sequence of the pCDVI-PL. Poly(dT)-tailed vector primer of pCDVI-PL was constructed by *Kpn*I digestion, and tailing with terminal nucleotidyl transferase, followed by cleavage with *Eco*RI. Using the pCDVI-PL primer vector and the dG-tailed linker containing the SP6 promoter described above, the cDNA library was constructed according to the Okayama-Berg method¹⁶.

Poly(A)⁺ mRNA of the 2.19 T-cell line was fractionated by centrifugation in a sucrose gradient. An aliquot of each fraction was injected into *Xenopus* oocytes and the incubation medium of the oocytes assayed for the activity of the IgG1 induction factor. Translation products directed by mRNAs in fractions 7-10 corresponding to 7S to 11S contained the IgG1 induction

activity with the peak activity at fraction 9 (data not shown). The result suggests that mRNA for the IgG1 induction factor may be ~700–900 base pairs (bp).

As the continuous culture of the 2.19 T-cell line required the presence of interleukin-2 (IL-2) and irradiated C57BL.10 spleen cells as stimulators, it was difficult to obtain $>10^9$ cells. The concentration of mRNA obtained was not enough to conduct the hybrid selection method for screening clones encoding the IgG1 induction factor. Therefore, we had to use an expression vector system that required only a limited amount of mRNA. The first vector we used was pCDV1 (ref. 16) which contained the SV40 promoter sequence. We constructed libraries of cDNA complementary to total poly(A)⁺ RNA and fraction 9 mRNA of the 2.19 T-cell line using the pCDV1 vector. DNAs of these libraries were used for transfection of COS 1 and COS 7 cells by the calcium phosphate precipitation method. We were unable to detect any IgG1 induction activity in culture supernatants of either COS 1 or COS 7 cells transfected with these library DNAs. This could have partly resulted from contamination of COS cells by mycoplasma.

Construction of pSP6K libraries

We needed a better expression vector to screen the cDNA library for the IgG1 induction factor. Some of us (Y.N., T.N., C.A., T.T., T.K., F.M., Y.Y. & T.H.) developed a new expression vector containing the SP6 promoter which allowed synthesis of a few micrograms of RNA by *in vitro* transcription with a specific RNA polymerase^{18,19}. A basic strategy of cDNA library construction is similar to that described by Okayama and Berg¹⁶ except that the *Hind*III linker segment containing the SV40 promoter sequence is replaced by another *Hind*III linker containing the SP6 promoter sequence. To obtain the SP6 promoter *Hind*III linker, the plasmid pSP62-PL was modified to pSP62-K2. The preparation of the linker and construction of the cDNA library are shown schematically in Fig. 1. We refer to this vector as the pSP6K system.

The pSP6K vector has several advantages over the pCDV1 vector and overcomes almost all the problems described above. Only a small amount (1–3 µg) of pSP6K DNA is required to be transcribed into mRNA, which is directly injected into *Xenopus* oocyte. This procedure avoids all the problems associated with DNA transfection to COS cells. Usually the amount of mRNA obtained by *in vitro* transcription with SP6 RNA polymerase is equal to that of template DNA. The pSP6K system is able to produce more concentrated supernatants than the pCDV1 system. Using cloned human IL-2 cDNA inserted into the pSP6K vector, we were able to obtain 1.6×10^5 U ml⁻¹ of the IL-2 activity in oocyte culture media. On the other hand, murine IL-2 cDNA in the pCDV1 vector produced 6.9×10^3 U ml⁻¹ of the IL-2 activity in culture supernatants of COS cells transfected with this clone²². Supernatants of *Xenopus* oocyte culture are free from mycoplasma contamination and contain few other proteins. We could concentrate oocyte supernatants a further 10-fold. In fact, previous cloning with the pCDV1 vector was carried out by pooling 40–50 clones, but we could easily detect activities of the IgG1 induction factor by pooling the total library (45,000 clones) for poly(A)⁺ mRNA (see below).

We have constructed two pSP6K libraries of cDNA; one for total poly(A)⁺ RNA from the 2.19 T-cell line and the other for fraction 9 RNA of the sucrose gradient fractionation. The former and the latter contained 4.5×10^4 and 4×10^3 independent cDNA clones, respectively. DNA of each library was cleaved with several restriction enzymes to linearize plasmid DNA, and capped RNA was synthesized using SP6 RNA polymerase. The synthesized RNA were injected into *Xenopus* oocytes and translation products secreted into the incubation media were assayed for the IgG1 induction factor. As shown in Table 1, the libraries for both total poly(A)⁺ RNA and fraction 9 mRNA significantly augmented the IgG1 response when template DNA was cleaved

Table 1 IgG1-inducing activities directed by pSP6K cDNA clones

Clones used as template	Restriction enzymes used for cleavage	IgG1 plaque-forming cells per culture
a		
Total cDNA library	<i>Sal</i> I	558
Total cDNA library	<i>Sac</i> I	12
Total cDNA library	<i>Pst</i> I	18
Phosphate buffer	None	138
Positive control	None	858
b		
Fraction 9 cDNA library	<i>Sal</i> I	490
Fraction 9 cDNA library	<i>Sac</i> I	53
Fraction 9 cDNA library	<i>Pst</i> I	17
Phosphate buffer	None	43
None	None	17
Positive control	None	1,516
c		
No. 280	<i>Sal</i> I	140
No. 293	<i>Sal</i> I	3,164
No. 294	<i>Sal</i> I	315
No. 359	<i>Sal</i> I	252
No. 362	<i>Sal</i> I	224
No. 371	<i>Sal</i> I	273
No. 374	<i>Sal</i> I	1,680
Phosphate buffer	None	84
None	None	413
Positive control	None	2,520

a, The total cDNA library was a mixture of 45,000 pSP6K clones containing cDNA complementary to total poly(A)⁺ RNA of 2.19 T cells; **b**, fraction 9 cDNA library was a mixture of 4,000 pSP6K clones containing cDNA complementary to fraction 9 mRNA of sucrose gradient fractionation; **c**, seven clones containing cDNA inserts longer than 500 bp were chosen among 50 members of two subpools (3–12 and 3–15). In experiments **a**, **b** and **c**, oocyte culture media were added to concentrations of 20, 20 and 2%, respectively, to the B-cell cultures with LPS. Positive control is the 2.19 T-cell culture supernatant, which was added to a concentration of 5%. Total RNA was extracted³⁷ from a Con A-activated B6.C-H-2(bm12) 2.19 T-cell line^{14,20} (10^9 cells). Poly(A)⁺ RNA (300 µg) was obtained by purification with an oligo(dT)-cellulose column. Poly(A)⁺ (200 µg) RNA was applied onto a 5–22% sucrose gradient, centrifuged at 36,000 r.p.m. for 15 h with a Hitachi RPS40 rotor and collected into 16 fractions. Aliquots (100–200 ng per 0.7 µl) of fractionated RNA were injected into 10 *Xenopus* oocytes. The IgG1 induction activity of oocyte incubation media was assayed as described below. cDNA libraries for poly(A)⁺ RNA and an active fraction (fraction 9) of the sucrose gradient were constructed using the SP6K vector system as described in Fig. 1 legend. Libraries of recombinant plasmids were introduced into *Escherichia coli*, and DNA were prepared by conventional procedures. A mixture of cDNA clone was digested with *Sac*I, *Sal*I or *Pst*I, endonuclease sites which exist in the polylinker region lying at the 3' end of the cDNA; and they are also found in the vector sequence, but not in the linker region containing the SP6 promoter. The digestion was required for termination of transcription. Digested DNA (3 µg) was dissolved in a 50-µl reaction mixture containing 40 mM Tris (pH 7.5), 6 mM MgCl₂, 2 mM spermidine, 0.5 mM nucleotide triphosphates, 0.1 mg ml⁻¹ bovine serum albumin, 10 mM dithiothreitol, 1 mM m⁷GpppG (Pharmacia), 50 U RNase inhibitor and 8 U SP6 RNA polymerase (NEN or Promega Biotec). For the estimation of the amount of the transcripts, ~20,000 c.p.m. [α -³²P]GTP or [α -³²P]UTP (3,000 Ci mmol⁻¹) was added. The reaction was carried out at 40 °C for 1 h as described elsewhere^{18,19,38}. The reaction mixture was passed through a Sephadex G-50 column, extracted with phenol and precipitated with ethanol. The RNA was dissolved with autoclaved water to a final concentration of 0.2–0.4 µg µl⁻¹. The average yield of transcripts from 3 µg template DNA was 2–4 µg. To examine lengths of transcripts, aliquots of transcripts were glyoxalated and analysed by agarose gel electrophoresis³⁹. Almost all the transcripts were full length. The capping efficiency using m⁷GpppG was almost the same as that using the capping enzyme¹⁹. Transcripts synthesized in the presence of m⁷GpppG were no longer a substrate for the capping enzyme. The transcripts (10–20 ng per 70 nl) were injected into one *Xenopus* oocyte. Ten oocytes for each sample of RNA were incubated in 100 µl Barth's medium at 20 °C for 36 h. Incubation medium was gently collected and aliquots were added to LPS-containing cultures of mouse spleen cells (200 µl per well) as described previously²⁰. C57BL.10(F₁) spleen cells (4×10^5 ml⁻¹) were cultured in RPMI 1640 containing L-glutamine, penicillin-streptomycin, 5×10^{-5} M 2-mercaptoethanol, 15% fetal calf serum (FCS) and 50 µg ml⁻¹ LPS and supernatants were added on day 0. The numbers of cells secreting immunoglobulin of different isotypes were estimated in the protein A plaque assay⁴⁰, using subclass-specific developing antisera¹⁴. The responses shown are from days 5 and 6.

The insert cDNA of the two clones were isolated by *Bam*HI digestion. The restriction endonuclease cleavage maps of the inserts were constructed by conventional procedures and summarized in Fig. 2a. As the maps of the two clones are indistinguishable from each other, we chose to use pSP6KmIL4-374 in the following experiments. Both pSP6KmIL4-293 and pSP6KmIL4-374 have *Pst*I and *Sac*I sites in the inserts, as expected from the fact that *Pst*I or *Sac*I digestion of the cDNA libraries destroyed the IgG1-inducing activity (Table 1).

The deduced amino-acid sequence of the IgG1 induction factor was compared with sequences of known proteins including all the lymphokines so far identified at the sequence level. A computer-assisted search revealed that segments of murine

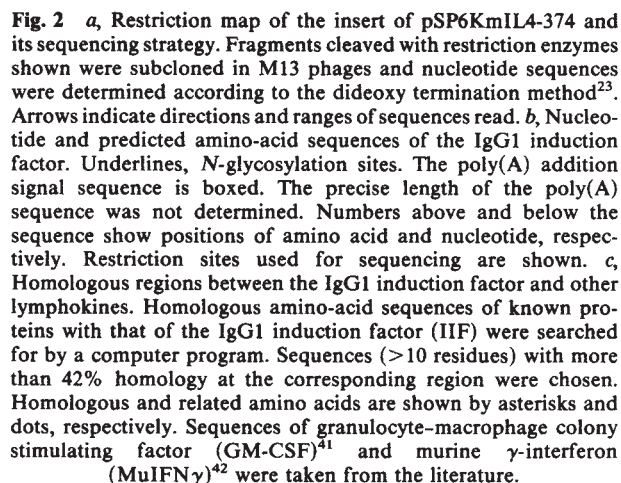


Fig. 2 *a*, Restriction map of the insert of pSP6KmIL4-374 and its sequencing strategy. Fragments cleaved with restriction enzymes shown were subcloned in M13 phages and nucleotide sequences were determined according to the dideoxy termination method². Arrows indicate directions and ranges of sequences read. *b*, Nucleotide and predicted amino-acid sequences of the IgG1 induction factor. Underlines, *N*-glycosylation sites. The poly(A) addition signal sequence is boxed. The precise length of the poly(A) sequence was not determined. Numbers above and below the sequence show positions of amino acid and nucleotide, respectively. Restriction sites used for sequencing are shown. *c*, Homologous regions between the IgG1 induction factor and other lymphokines. Homologous amino-acid sequences of known proteins with that of the IgG1 induction factor (IIF) were searched for by a computer program. Sequences (>10 residues) with more than 42% homology at the corresponding region were chosen. Homologous and related amino acids are shown by asterisks and dots, respectively. Sequences of granulocyte-macrophage colony stimulating factor (GM-CSF)⁴¹ and murine γ -interferon (MuIFN γ)⁴² were taken from the literature.

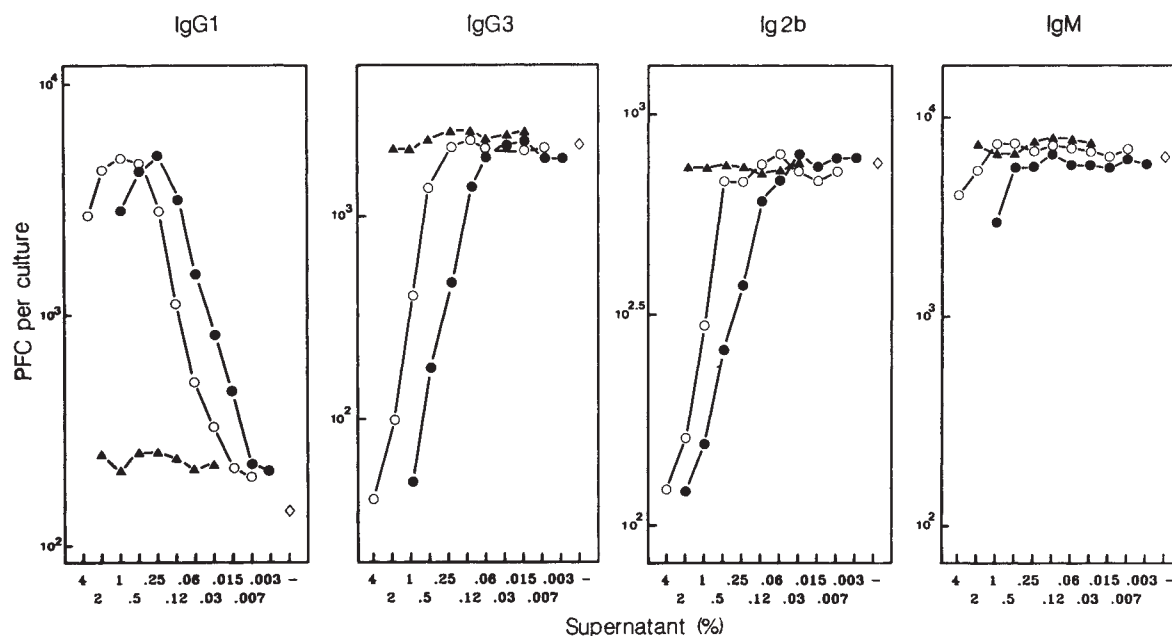


Fig. 3 Effect of translation products of pSP6KmIL4-374 on the isotypic profile of LPS-stimulated spleen cell cultures. The samples added at indicated concentrations were: ●, supernatants from oocytes injected with mRNA encoded by the pSP6KmIL4-374 cDNA clone (positive cDNA clone); ▲, supernatants from oocytes injected with mRNA encoded by the pSP6K280 cDNA clone (negative cDNA clone); ○, supernatants from the 2.19 cell line activated with ConA (positive control); ◇, LPS response in the absence of additional supernatants. Preparation of the supernatants and plaque-forming cell (PFC) assay were done as described in Table 1 legend.

granulocyte-macrophage colony stimulating factor (GM-CSF) and murine γ -interferon contain significant homologies with the IgG1 induction factor (Fig. 2c). Two regions (residues 67-78 and 115-133) contain five and eight matched residues, respectively, at corresponding regions of GM-CSF. One region (residues 72-81), which overlapped partially with the more N-proximal of the above two, contained six matched residues with a corresponding region of γ -interferon. If we assume that related amino acids such as Ser/Thr, Leu/Ile, Glu/Asp and Arg/Lys are homologous, the similarity increases to 63-68%. Although similar extents of homology were found in segments of other genes, only GM-CSF and γ -interferon contained homologous regions at the equivalent positions within the molecules. The results suggest that the IgG1 induction factor is distantly related to these lymphokines.

Biological activities

The cDNA for the IgG1 induction factor was the first B-cell factor to be cloned. We have further tested this lymphokine for several other activities that are thought to be associated with IgG1 induction factor. First, we asked whether the product of a single gene can induce IgG1 production and reduce IgG3 and IgG2b secretion. Supernatants of oocytes injected with pSP6KmIL4-374 mRNA or from the 2.19 T-cell line were added to LPS cultures and the IgG subclass responses were assessed (Fig. 3). In the absence of added supernatants, LPS stimulated high IgM, IgG2b and IgG3 responses, but a low IgG1 response¹⁴. Addition of either of the supernatants resulted in a dose-dependent increase in IgG1 secretion and a concomitant decrease in the level of IgG2b and IgG3 isotypes. No significant effect by

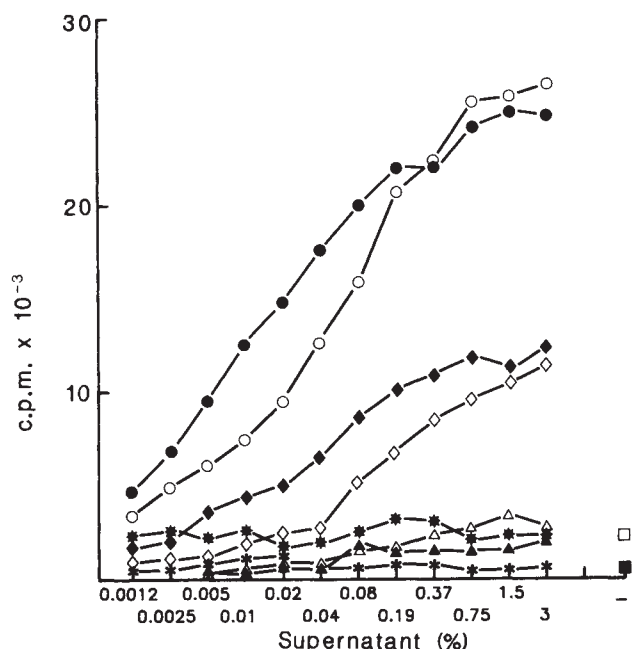
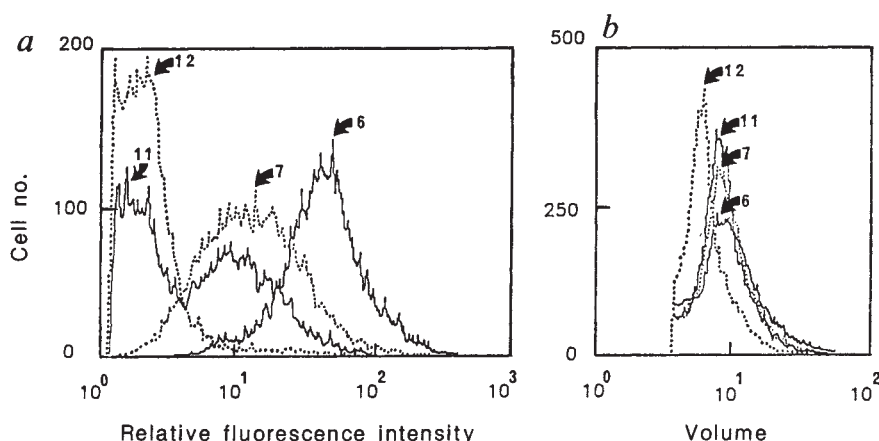


Fig. 4 Supernatants from the 2.19 T-cell line or from oocytes injected with pSP6.K374 mRNA are active in the BSF-1 assay. Indicated concentration of 2.19 T-cell supernatants were added to B-cell cultures either alone (Δ), together with 2 (◇) or 10 (○) $\mu\text{g ml}^{-1}$ F(ab')_2 anti-immunoglobulin. Translation products of pSP6KmIL4-374 were added alone (▲), together with 2 (●) or 10 (●) $\mu\text{g ml}^{-1}$ F(ab')_2 anti-immunoglobulin. Supernatants from oocytes injected with phosphate buffer were added together with 2 (*) or 10 (*) $\mu\text{g ml}^{-1}$ F(ab')_2 anti-immunoglobulin. Responses of cultures receiving F(ab')_2 anti-immunoglobulin alone at 2 (■) or 10 (□) $\mu\text{g ml}^{-1}$ are also shown.

Methods. B-cell enriched populations were prepared by running erythrocyte-free spleen cells from C3H/Tif mice through Sephadex G10 columns and thereafter treating them with monoclonal anti-Thy 1.2 antibodies and complement. This procedure resulted in populations containing >95% Ig^+ cells, which hereafter are called B lymphocytes. B cells (5×10^5) were cultured in the presence or absence of F(ab')_2 fragments of rabbit anti-mouse immunoglobulin antiserum and different concentrations of supernatants from 2.19 or oocytes. One day later the cultures were pulsed with ^3H -thymidine and collected on day 2. Acid-insoluble radioactivities were determined.

Fig. 5 Induction of increased expression of Ia by supernatants from the 2.19 T-cell line or oocytes injected with pSP6Kml4-374 mRNA. B cells ($4 \times 10^5 \text{ ml}^{-1}$) were cultured for 12 h in medium plus 5% FCS in 10-ml cultures with the indicated supplements. *a*, Cells were thereafter washed and stained with fluorescein-conjugated monoclonal anti-Ia (H118-49)⁴³ and analysed for 12 h in medium plus 5% FCS in 10-ml cultures with the indicated supplements. *a*, Cells were thereafter washed and stained with fluorescein-conjugated monoclonal anti-Ia (H118-49)⁴³ and analysed by flow cytometry using forward narrow-angle light scatter to facilitate exclusion of dead cells and log of integrated fluorescence to determine relative marker expression. 10,000 cells were analysed for each sample. Representative fluorescence (*a*) and forward narrow-angle light scatter profiles (*b*) are shown. Numbers refer to the experiments in Table 2.



the supernatants on the IgM response was observed. Supernatants derived from oocytes injected with mRNA from a negative clone (pSP6K280) had no effect on any isotype. These data demonstrate that the changes in the IgG1, IgG2b and IgG3 responses are caused by the products of a single gene, as suggested previously²⁰.

Based on the biochemical similarities between the IgG1 induction factor and the B-cell stimulating factor 1 (BSF-1)²⁷, some of us have previously suggested that the two factors are identical²⁸. The latter factor is defined by its capacity to act as a co-stimulator of DNA synthesis together with anti-immunoglobulin. As shown in Fig. 4, supernatants from the 2.19 T-cell line or translation products of the cDNA clone pSP6Kml4-374 were able to synergize with anti-immunoglobulin in inducing DNA synthesis in B lymphocytes. Supernatants from oocytes injected with phosphate buffer were inactive in this assay.

Another function that has been assigned to BSF-1 is its capacity to induce Ia expression in resting B lymphocytes²⁹. We show here that the product of the cloned gene as well as supernatants from the 2.19 T-cell line induced an almost three-fold increase in the amount of Ia expressed on B lymphocytes without any significant increase in cell size (Table 2, Fig. 5).

These data, together with the fact that the two molecules have the same biochemical characteristics and that affinity-purified BSF-1 has IgG1-inducing as well as IgG2b- and IgG3-reducing activities (ref. 28 and E. S. Vitetta and W. E. Paul, personal communication), show conclusively that the IgG1 induction factor and the BSF-1 are the products of the same gene.

Because of the complexity of the process of B-cell activation, the different steps involved have previously been considered to be independently controlled events. Several experimental systems meant to identify and to characterize the parameters regulating each of these events were constructed. As a consequence many B-cell-specific factors have been reported and classified as growth, differentiation or maturation factors^{12,15,30-35}. The fact that a single molecular species can participate in several activation processes has implications for our view of B-cell activation. It may be an oversimplified view of this dynamic process to consider the B-cell response as an ordered series of independent events, each of them regulated independently by the action of different controlling elements.

As the same molecule is active in so many biological assays, the names B-cell growth factor-1, BSF-1, BCDF γ and IgG1 induction factor that have previously been assigned to various activities of this lymphokine seem inadequate to describe this molecule defined by the structure shown in Fig. 2b. This lymphokine fulfills the necessary criteria to be termed an interleukin. Thus, we propose that this lymphokine should be called interleukin-4.

We thank Professor T. Ohi for advice about the homology search; Ms Y. Utsunomiya and L. Gerggren for technical assistance; Dr G. Möller for reading the manuscript; and Ms S.

Table 2 Effect of 2.19 T-cell line or oocytes injected with pSP6Kml4-374 mRNA on Ia expression

Cells analysed	Treatment	Mean fluorescence channel
1 B cells	—	21
2 B cells	1% 2.19 supernatant	53
3 B cells	0.3% 2.19 supernatant	47
4 B cells	0.6% pSP6.K374	51
5 B cells	0.2% pSP6.K374	48
6 B cells	0.07% pSP6.K374	52
7 B cells	0.6% PBS-injected oocyte supernatant	23
8 B cells	50 $\mu\text{g ml}^{-1}$ LPS	31
9 B cells	1 $\mu\text{g ml}^{-1}$ F(ab') ₂ anti-Ig	30
10 B cells	10 $\mu\text{g ml}^{-1}$ F(ab') ₂ anti-Ig	37
11 B cells in normal spleen	—	18
12 Thymocytes	—	3

See Fig. 5 legend for a description of the experiments.

Okamoto for preparation of the manuscript. This investigation was supported in part by grants from the Ministry of Education, Science and Culture of Japan, the Swedish National Science Research Council, the Swedish MRC and the Aristotelis Onassis public benefit foundation.

Note added in proof: Since submission of this manuscript we learned that cDNA for the same lymphokine has been cloned by F. Lee *et al.* (*Proc. natn. Acad. Sci. U.S.A.*, in the press), who have also agreed to call this lymphokine interleukin-4. We have more recently learned that C. J. Sanderson *et al.* (*Proc. natn. Acad. Sci. U.S.A.*, in the press) have proposed the name interleukin-4 for eosinophil differentiation factor.

Received 4 October; accepted 12 December 1985.

- Honjo, T. & Kataoka, T. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2140-2144 (1978).
- Kataoka, T., Kawakami, T., Takahashi, N. & Honjo, T. *Proc. natn. Acad. U.S.A.* **77**, 919-923 (1980).
- Davis, M. M. *et al.* *Nature* **283**, 733-739 (1980).
- Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286**, 676-683 (1980).
- Kataoka, T., Miyata, T. & Honjo, T. *Cell* **23**, 357-368 (1981).
- Yaoita, Y. & Honjo, T. *Nature* **286**, 850-853 (1980).
- Nikaido, T., Yamawaki-Kataoka, Y. & Honjo, T. *J. biol. Chem.* **257**, 7322-7329 (1982).
- Perlmutter, R. M., Hansburg, D., Briles, D., Nicolotti, R. A. & Davie, J. M. *J. Immun.* **121**, 566-572 (1978).
- Slack, J. N., Der-Balian, G. P., Nahm, M. & Davie, J. M. *J. exp. Med.* **151**, 853-862 (1980).
- Izui, S., Eisenberg, R. A. & Dixon, F. J. *J. exp. Med.* **153**, 324-338 (1981).
- Cebra, J. J., Komisar, J. I. & Schweitzer, P. A. *Rev. Immun.* **2**, 493-548 (1984).
- Isakson, P. C., Pure, E., Vitteta, E. S. & Krammer, P. H. *J. exp. Med.* **155**, 734-748 (1982).
- Coutinho, A., Pettersson, S., Ruuth, E. & Forni, L. *Eur. J. Immun.* **13**, 269-272 (1983).
- Bergstedt-Lindquist, S., Sideras, P., MacDonald, H. R. & Severinson, E. *Immun. Rev.* **78**, 25-50 (1984).
- Kishimoto, T. A. *Rev. Immun.* **3**, 133-158 (1985).
- Okayama, H. & Berg, P. *Molec. cell. Biol.* **3**, 280-289 (1983).
- Yokota, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 1070-1074 (1984).
- Melton, D. A. *et al.* *Nucleic Acids Res.* **12**, 7035-7056 (1984).
- Krieg, P. A. & Melton, D. A. *Nucleic Acids Res.* **12**, 7057-7070 (1984).
- Sideras, P., Bergstedt-Lindquist, S., MacDonald, H. R. & Severinson, E. *Eur. J. Immun.* **15**, 586-593 (1985).

21. Gurdon, J. B., Lane, C. D. & Woodland, H. R. *Nature* **233**, 177-184 (1971).
22. Yokota, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 68-72 (1985).
23. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5469 (1977).
24. Altenburger, W., Steinmetz, M. & Zachau, H. G. *Nature* **287**, 603-607 (1980).
25. Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. *Cell* **27**, 573-581 (1981).
26. Rechavi, *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 4405-4409 (1982).
27. Howard, M. & Paul, W. E. *A. Rev. Immunol.* **1**, 307-333 (1983).
28. Sideras, P., Bergstedt-Lindqvist, S. & Severinson, E. *Eur. J. Immun.* **15**, 593-598 (1985).
29. Noelle, R., Krammer, P. H., Ohara, J., Uhr, J. W. & Vitetta, E. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6149-6153 (1984).
30. Möller, G. (ed.) *Immun. Rev.* **78** (1984).
31. Schimpl, A. & Wecker, E. *Transplant. Rev.* **23**, 176-188 (1975).
32. Swain, S. & Dutton, R. J. *exp. Med.* **156**, 1821-1834 (1982).
33. Howard, M. *et al. J. exp. Med.* **155**, 914-923 (1982).
34. Anderson, J. & Melchers, F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2497-2501 (1981).
35. Leanderson, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7455-7459 (1982).
36. Aarden, L. A. *et al. J. Immun.* **123**, 2928-2929 (1979).
37. Ullrich, A. *et al. Science* **196**, 1313-1319 (1977).
38. Konarska, M. M., Padgett, R. A. & Sharp, P. A. *Cell* **38**, 731-736 (1984).
39. Makmaster, G. K. & Carmichael, G. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4835-4838 (1977).
40. Gronowicz, E., Couthino, A. & Melchers, F. *Eur. J. Immun.* **6**, 588-590 (1976).
41. Gough, N. M., Metcalf, D., Gough, J., Grail, D. & Dunn, A. R. *EMBO J.* **4**, 645-653 (1985).
42. Gray, P. W. & Goeddel, D. V. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5842-5846 (1983).
43. Lemke, H., Hammerling, G. J. & Hammerling, U. *Immun. Rev.* **47**, 175-206 (1979).

LETTERS TO NATURE

Is IRAS cirrus cloud emission largely fine-structure radiation?

Martin Harwit, James R. Houck & Gordon J. Stacey*

Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA

One of the more interesting observations made by the infrared astronomical satellite (IRAS) mission is the discovery of a diffuse 'cirrus' component seen at 60 and 100 μm (bands 3 and 4). It straddles the galactic plane and extends, in patches, to latitudes of $\pm 80^\circ$ (refs 1, 2). Comparison of the observed brightness from these clouds with [O I] 63- μm and [O III] 88- μm fine-structure line emission, calculated on the basis of observed [C II] emission from the galactic plane, shows the flux levels observed by IRAS to be of the same order as the expected fine-structure line emission³. We propose here that fine-structure line emission, from neutral gas at 63 μm and from diffuse ionized clouds at 88 μm , is a significant component and in some clouds perhaps the prime source of the cirrus radiation.

Hauser *et al.*² have reported on the diffuse far-infrared flux from the galactic poles, and Low *et al.*¹ cite the emission from four different regions at high galactic latitudes. Table 1 summarizes these results. Low *et al.* add the following relevant comments: "There are regions of relatively prominent 100 μm emission with only weak H I, and there are many structures in the H I maps that do not appear prominent in the 100 μm scans." Similarly, a comparison of cirrus cloud emission⁴ and maps of dust polarization at high galactic latitudes⁵ shows significant differences.

If cirrus clouds had to be modelled solely in terms of the emission from dusty regions, the ratio of 60- to 100- μm flux, $F(60)/F(100)$ listed in Table 1, would also be difficult to explain. Spencer and Leung⁶ would argue that this ratio should be 0.55 for graphite grains and 0.03 for silicate grains illuminated by a homogeneous interstellar radiation field¹. On this basis, cirrus clouds would have to contain a substantial fraction of graphite grains, but there is little or no observational evidence for this. We note, however, that Draine and Anderson avoid this difficulty by introducing the assumption of mixtures of graphite and silicate grains with size distributions reaching down to 3- $\text{\AA}$ radius⁷.

Table 1 also suggests that different observed clouds contain different mixtures of grains, although this conclusion might be challenged by observations reported by Boulanger *et al.*⁸. They find a steadier correlation for the 60- μm flux with the 100- μm flux as well as with the H I column density over a $20^\circ \times 18^\circ$ field they investigated around RA = 14 h 40 min, dec. = 48° .

Nevertheless, there are several difficulties with models of cirrus cloud emission which restrict themselves to dust emission. It is therefore worth examining the extent to which atomic and ionic fine-structure emission could contribute to the cirrus radiation.

The strongest lines from the neutral hydrogen component of the interstellar medium are the 157- μm line of singly ionized carbon and the 63- μm line of neutral oxygen. In low-density ionized regions, the 88- and 52- μm emission lines caused by doubly ionized oxygen are strong. Of these lines, the 52- and 63- μm emission would contribute to the IRAS 60- μm band and the 88- μm emission would contribute mainly to the 100- μm band.

Recent observations³, integrated over the entire galactic disk, show that the total C⁺ fine-structure luminosity is roughly $6 \times 10^7 L_\odot$. If this radiation were to be uniformly emitted from the entire disk, $\sim 10^{46} \text{ cm}^2$ in area, we would expect a flux of $6 \times 10^{-13} \text{ W cm}^{-2} \text{ sr}^{-1}$ from each pole. The C⁺ fine-structure emission seems to originate at the interface between ionized and neutral hydrogen regions. Calculations, based on the total amount of ionized and neutral gas in the Galaxy, as well as on gas observed to lie in interfacial regions, suggest that the 63-, 88- and 157- μm emission ought to be comparably strong in the Galaxy⁹. In particular, Stacey calculates that the 88- μm [O III] fine-structure emission from extended low-density ionized gas in the Galaxy should amount to somewhat more than $2 \times 10^7 L_\odot$, just slightly higher than the 157- μm [C II] emission from the same regions. At H II/H I interfaces, the [C II] emission is about twice this value, and the 63- μm emission due to [O I] fine-structure transitions is expected to be $3 \times 10^7 L_\odot$. These line fluxes, when viewed by the broad-band IRAS detectors, would appear to have levels of the order of 0.1 MJy sr⁻¹. If the [O III] 52- μm and [N III] 57- μm lines were also strong, levels of the order of 0.3 MJy sr⁻¹ would be expected in band 3, comparable with values cited in Table 1.

The calculated ratio of galactic 63:88:157- μm emission cited above is based on models of neutral, ionized and interfacial regions in the Galaxy⁹. For line intensities of [O I] 63 μm , [O III] 88 μm , [C II] 157 μm , [O III] 52 μm and [N III] 57 μm , at the centre of the galaxy M82, approximate ratios of 5:5:5:2:1 have been reported, with somewhat larger fields of view used for observing the C II emission¹⁰⁻¹². This supports the hypothesis that substantial fractions of the cirrus emission may be due to fine-structure lines.

Another statistical indication comes from a correlation of [C II] and to a lesser extent [O I] emission with ¹²CO, $J = 1 \rightarrow 0$ line strength¹¹. All three types of emission appear to emanate predominantly from surfaces of molecular clouds illuminated by ultraviolet radiation¹³. Given that correlation, one would expect to find a correlation between ¹²CO emission and 63- μm [O I] line radiation, and therefore band 3 cirrus radiation at intermediate galactic latitudes.

In this connection, recent measurements by Magnani *et al.* are particularly relevant¹⁴. These observers examined 488 regions of the sky characterized either by high optical extinction

* Present address: Physics Department, University of California, Berkeley, California 94720, USA.

Table 1 Emission at high galactic latitudes

		60 μm	100 μm	$F(60)/F(100)$	A_v/τ^*	$\text{H I}^* (10^{20} \text{ cm}^{-2})$
South galactic pole		$\leq 0.7 \text{ MJy sr}^{-1}$	2.4 MJy sr^{-1}	≤ 0.3	—	—
North galactic pole		≤ 0.7	2.0	≤ 0.3	—	—
$l = 129.7^\circ$	$b = -69.4^\circ$	3.0	10	0.3	3,500	6.3
237.0	74.7	0.79	4.3	0.18	2,300	5.2
272.7	81.7	0.20	1.3	0.15	2,300	4.0
252.9	61.4	1.0	4.9	0.2	2,000	5.2

* H I column density and A_v values are from ref. 20. l is the galactic longitude, A_v the visual absorption in magnitudes and τ is the emission optical-depth at 100 μm .

or high optical surface brightness at substantial galactic latitudes. CO emission was observed from 57 distinct molecular clouds comprising 35 individual complexes at galactic latitudes $|b| > 25^\circ$. These clouds were correlated with H I concentrations, although molecular hydrogen appeared to predominate in most cases. Individual clouds subtend areas of fractions of square degrees ranging up to several square degrees, implying, the authors suggest, that they are responsible for at least some of the diffuse high-latitude infrared cirrus emission, despite the surprising fact that the particular clouds listed by Low *et al.*¹ all failed to exhibit CO emission.

If the general correlations noted by Crawford *et al.*¹¹ are applicable to high-galactic-latitude clouds, a [C II] flux of $5 \times 10^{-12} \text{ W cm}^{-2} \text{ sr}^{-1}$ would be expected for an integrated CO line intensity $I_{\text{CO}}^* = 4 \text{ K km s}^{-1}$, a flux comparable with that found in many of the Magnani *et al.* clouds. If the 63- μm emission were half as great as the [C II] flux, we would be seeing fluxes of the order of 1 MJy sr^{-1} from such CO-emitting clouds, again a value comparable with those cited by Low *et al.*

The mean radiation in all of these lines therefore approaches values measured by IRAS for typical high-galactic-latitude regions of the sky. If the sky is patchy, as is shown by the H I observations, individual regions should be emitting flux levels in excess of several MJy sr^{-1} —the actual levels observed in IRAS scans.

The extent to which line radiation contributes to the cirrus cloud emission could be tested most directly through far-infrared spectral observations. The diffuse nature of these clouds, however, makes such observations difficult. Stacey *et al.*³ have obtained observations for the [C II] 157- μm emission near the galactic plane using the NASA Lear jet telescope. Unfortunately, the Lear jet is currently not available for astronomical studies.

Although direct observation is impossible, there are several possible indirect approaches:

(1) Heiles¹⁵ has emphasized the existence of H I clouds which produce no optical extinction and therefore appear to have extremely low dust concentrations. Such regions would, on our hypothesis, nevertheless be expected to exhibit line emission and statistically should contribute to the cirrus emission, despite the lack of dust.

(2) Statistical correlations should also be expected between bright cirrus regions and clouds that are powerful emitters of radio recombination lines. For some of these one would expect contributions from the [O III] 88- and 52- μm lines; for others, contributions from the [O I] 63- μm lines. If both hydrogen and helium recombination lines are observed, the state of ionization of such clouds becomes apparent, and the extent to which the [O I] or [O III] lines should prevail could be estimated. In turn, that would lead to predictions of the contributions these lines would respectively make to the IRAS 60- and 100- μm bands.

An alternative optical approach would be to map diffuse H α , [O III] 5,007- $\text{\AA}$, and [O I] 6,300- $\text{\AA}$ emission at intermediate galactic latitudes, although the interpretation of any observed correlations with IRAS data could be ambiguous.

(3) Evidence of cirrus emission at short wavelengths, suggesting improbably high temperatures for high-galactic-latitude grain emission, would also indicate the possibility of shorter-wavelength emission from ionized gas through [Ne II] 12.8- μm or [S IV] 10.4- μm emission detectable in IRAS band 1, or [S III] emission at 18.7 μm detectable in band 2. Contributions from other fine-structure transitions at short wavelengths are also conceivable. The small grains of Draine and Anderson, however, do offer a way to account for the high emission at short wavelengths, since small grains can be briefly heated even by single-photon absorption^{16,17}. A similar explanation is offered by Puget *et al.*, who suggest admixtures of large polycyclic molecules; these had already been introduced by Leger and Puget to account for the unidentified spectral features seen in the interstellar emission at 3.3, 6.2, 7.8, 8.6 and 11.3 μm ^{18,19}.

(4) Evidence against fine-structure emission dominating at 100 μm would come from consistently identical ratios of $F(60)/F(100)$ and of $F(100)$ to the H I column density. According to our hypothesis, the 100- μm flux would probably be due to 88- μm line emission, which would not have a constant relation to atomic hydrogen column densities, although the 63- μm [O I] emission could.

We thank Dr Leo Blitz for permitting us to see his results before publication, and Drs Michael Hauser and Michael Rowan-Robinson for incisive comments. This work was supported by NASA grants NSG 2347 and NAGW-761 and JPL contract 954603.

Received 23 September; accepted 10 December 1985.

- Low, F. J. *et al. Astrophys. J.* **278**, L19–L22 (1984).
- Hauser, M. G. *et al. Astrophys. J.* **278**, L15–L18 (1984).
- Stacey, G. J., Viscuso, P. J., Fuller, C. E. & Kurtz, N. T. *Astrophys. J.* **289**, 803–806 (1985).
- Habing, H. J. & Neugebauer, G. *Scient. Am.* **251**(5), 48–57 (1984).
- Mathewson, D. S. & Ford, V. L. *Mem. R. astr. Soc.* **74**, 139–182 (1970).
- Spencer, R. G. & Leung, C. M. *Astrophys. J.* **222**, 140–152 (1978).
- Draine, B. T. & Anderson, N. *Astrophys. J.* **292**, 494–499 (1985).
- Boulanger, F., Baud, B. & van Albada, G. D. *Astr. Astrophys.* **144**, L9–L12 (1985).
- Stacey, G. J. thesis, Cornell Univ. (1984).
- Watson, D. M., Genzel, R., Townes, C. H., Werner, M. W. & Storey, J. W. V. *Astrophys. J.* **279**, L1–L4 (1984).
- Crawford, M. K., Genzel, R., Townes, C. H. & Watson, D. M. *Astrophys. J.* **291**, 755–771 (1985).
- Duffy, P., Erickson, E., Haas, M. R. & Houck, J. R. preprint, NASA, Ames Research Center (1985).
- Genzel, R. & Stacey, G. L. *Mitt. d. Astr. Gesellschaft* (in the press) (1985).
- Magnani, L., Blitz, L. & Mundy, L. preprint, University of Maryland (1985).
- Heiles, C. preprint, University of California, Berkeley (1985).
- Harwit, M. *Astrophys. J.* **199**, 398–400 (1975).
- Sellgren, K. *Astrophys. J.* **277**, 623–633 (1984).
- Puget, J. L., Léger, A. & Boulanger, F. *Astr. Astrophys.* **142**, L19–L22 (1985).
- Léger, A. & Puget, J. L. *Astr. Astrophys.* **137**, L5–L8 (1984).
- Burstein, D. & Heiles, C. *Astr. J.* **87**, 1165–1189 (1982).

Quasi-periodic oscillations from accretion disk coronas

C. B. Boyle, A. C. Fabian & P. W. Guilbert

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

Recent observations of galactic bulge X-ray sources have revealed in some cases the existence of a hot accretion disk corona^{1,2} and in others quasi-periodic oscillations (QPOs) in the X-ray flux³⁻⁵. We suggest here that these properties are directly related and that the oscillations are produced in the innermost parts of the corona. The observed variations in the centroid frequency of the QPOs with source intensity³⁻⁶ are explained and further variations with the photon spectrum predicted. Unlike some alternative models^{7,8}, we do not require that the central object is a rapidly spinning weakly magnetized neutron star.

Attention has recently focused on the discovery of intensity-dependent QPOs in the X-ray flux from several galactic bulge sources, in particular GX5-1 (ref. 3), Sco X-1 (ref. 5) and Cyg X-2 (ref. 4). The centroid frequency of these oscillations varied between 20 and 40 Hz in the case of GX5-1 (ref. 3), 5 and 10 Hz in the case of Sco X-1 (ref. 5) and 35 to 40 Hz in Cyg X-2 (ref. 4). In all three sources, this frequency is observed to increase with increasing intensity; the range of the QPO periods also shows a correlation with the intensity in the case of GX5-1 but not so clearly in the other two sources. Another characteristic is that the QPOs are observed only when the sources are quiescent, being totally absent during flaring. There is also some evidence for correlation with the changing spectral state of the source⁶, although observations are not conclusive. QPOs have also been observed⁹⁻¹¹ from the candidate black-hole binary GX339-4 at frequencies ~ 0.1 Hz.

The observations of partial X-ray eclipses in the sources 4U1822-37 (ref. 1) and 4U2129+47 (ref. 2) and the residual emission from Her X-1 during its low states¹² have led to the suggestion^{13,14} that galactic bulge X-ray sources have a corona of hot gas ($T > 10^7$ K) above their accretion disks with appreciable scattering optical depths to the observer. X-ray heating of the upper layers of the accretion disk surrounding a compact object can result in the formation of an optically thick corona at the Compton temperature of the incident X rays.

We suggest that a connection exists between the phenomenon of QPOs and accretion disk coronas (ADC). In particular, the frequencies of the QPOs are identified with those of oscillations of disturbances in the inner regions of a corona above an accretion disk. Such oscillations will occur predominantly at ν_k , the keplerian angular frequency of disk matter about the central object at the radius of interest and have low coherence (ref. 15). Accretion is observed to be a generally noisy process (compare Cyg X-1) and can provide the power to excite such oscillations.

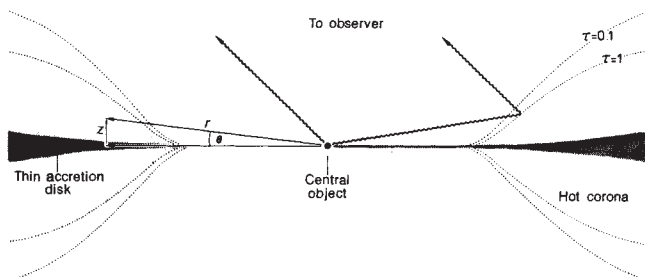


Fig. 1 Representation of the coronal model presented here. The point (r, θ) is any general point in the hot corona. Wavy lines represent X-ray photons from the compact central source.

Consider a situation in which a fraction up to $\sim 10\%$ of the observed X rays are scattered from the $\tau = 1$ surface of the inner corona (τ being the electron scattering optical depth to the source; see Fig. 1). Identifiable QPOs will occur when this surface rises sufficiently steeply with radius that the scattered X rays come from a restricted range of radii and hence frequency. The amplitude and coherence of such oscillations would then depend on the shape of the $\tau = 1$ surface, the radial gradient of τ at this surface, the height of the corona (as seen by the central source) and on the number of independently oscillating 'blobs' per unit circumference.

An isothermal Compton-heated corona at temperature T_c equal to the radiation temperature of the source in hydrostatic equilibrium results in a gaussian density profile which, in terms of the angle $\theta = z/r$ (see Fig. 1) is given by

$$n(r, \theta) = n_0(r) \exp(-\beta \theta^2/r) \quad (1)$$

(assuming θ is small). The parameter β is given by

$$\beta = \frac{GM_x \mu m_H}{4 k T_c} \quad (2)$$

With M_x the mass of the accreting object and μ the molecular weight of the coronal gas. Above a cold thin disk, the density at the base of the corona, $n_0(r)$, will be determined by the balance of ionization and recombination (assuming that Compton heating occurs at an infinitesimal height above the disk). Thus, $n_0 = F_x/\xi$, F_x being the ionizing X-ray flux incident on the disk and ξ being the ionization parameter ($\approx 10^3$ for total ionization of gas with solar abundances).

Assuming for the present that a constant fraction, f , of the incident flux, L/r^2 , where L is the luminosity of the central source, gets scattered down onto the disk, the electron scattering optical depth along a direction θ to a distance r from the compact object is

$$\tau = \frac{\alpha L \sigma_T}{\beta \theta^2} \exp(-\beta \theta^2/r) \quad (\theta \ll 1) \quad (3)$$

with $\alpha = 10^{-3} f$ and σ_T the Thomson cross-section for electron scattering. Taking logarithms and replacing r by the keplerian frequency ν , gives

$$\nu = \nu_0 (\log_{10} I - C_0)^{3/2} \quad (4)$$

where I is the intensity in counts s^{-1} , with

$$C_0 = \log_{10} \frac{\tau \beta \theta^2}{\sigma_T \alpha} + \log_{10} \frac{I}{L} \quad (5)$$

and

$$\nu_0 = \frac{1}{2\pi} \left(\frac{2.3(GM_x)^{1/3}}{\beta \theta^2} \right)^{3/2} \quad (6)$$

We have compared equations (4)-(6) with observed values of QPO frequency and intensity for GX5-1 (ref. 3). The fit appears to be as good as the expressions quoted by van der Klis *et al.*³ ($\nu \propto I^2$ and $\nu = \text{constant} \times I^{3/7} - \nu_0$), although at the highest intensities the quadratic function is a little inadequate. At these large intensities, however, the centroid frequency of the QPOs becomes difficult to measure. As the variation in frequency has only been measured over a small range of intensity, almost any two-parameter model would give an adequate fit. The situation is no better in the case of the other two sources.

Fitting equations (4)-(6) to the GX5-1 data yields $C_0 = 3.1$ and $\nu_0 = 144$ Hz. This value for ν_0 implies $T_c/\theta^2 \approx 3 \times 10^{10}$ which for a temperature of 5×10^7 K implies $\theta \approx 2^\circ$. Assuming a luminosity of 10^{38} erg s^{-1} , then a count rate of 3,000 counts s^{-1} and the above value of C_0 imply $\tau/\alpha = 1.5 \times 10^8$ which for $\tau = 1$ means $f \approx 10^{-4}$. These values with $M_x = 1.4 M_\odot$ ($\mu = 1$) give $r \approx 2.3 \times 10^7$ cm or a frequency of 32 Hz, consistent with the typical frequency observed.

Because in this model $\nu \propto T_c^{3/2} (\ln T_c)^{3/2}$, spectral changes in the source will complicate the behaviour of the frequency-

intensity correlation. For example, a spectral softening of the source with increasing intensity would lead to an anticorrelation of frequency and intensity, the form depending on the relationship between intrinsic luminosity and Compton temperature. It also predicts that in the absence of any nonlinear dependence on intensity, the frequency is more sensitive to the source temperature than to the intensity and so changes in frequency may occur in situations where the source intensity is constant but there is a change in T_c . In the case of Cyg X-2, Branduardi-Raymont *et al.*¹⁶ have found a nonlinear dependence of spectral hardness on intensity, which means that the source is either soft or hard at low intensities. This behaviour maps through T_c onto the (ν, I) plane and explains why Cyg X-2 does not always oscillate at the same frequency for a given intensity. However, one must be careful to separate out intrinsic changes in the source from observed changes, as scattering in the hot corona will greatly affect the spectral behaviour. Whether the photon spectrum of the QPOs should be softer or harder than the overall spectrum depends on the average number of scatterings and the direct contribution by the central source. The QPOs should be softer if most of the other photons undergo less than one scattering, and harder if they scatter more. The QPOs should also be more polarized. Sources with intrinsically soft spectra should show low QPO frequency. Thus, a factor of 3 decrease in T_c for GX5-1 would cause the QPO to occur around 6 Hz, as observed for Sco X-1 (ref. 2). In its simplest form, our model predicts that all QPO sources lie on the same plane in (ν, L, T_c) space. The disk inclination will influence this.

The absence of QPOs during flaring activity may be due to a 'puffing up' of the corona with increasing intensity, which would drastically decrease the coherence of any oscillations due to a change in the shape and position of the $\tau=1$ surface and the amount of obscuration along our line-of-sight to the inner corona. We rely on accretion noise to excite the oscillations in the corona picked out as QPO at the $\tau=1$ surface. If power spectra of Cyg X-1 are a guide, we expect that accretion noise will lead to excess low-frequency noise in the power spectrum of QPO sources, which should generally correlate with the QPO power.

From equation (1) the change in QPO frequency with optical depth, which might be regarded as strongly determining the period-range of the oscillations, is a slowly varying function of depth:

$$\frac{d\nu}{d\tau} = \frac{-1}{\tau} (\log_e (L/\tau))^{1/2} \quad (7)$$

This predicts that the range of QPO periods will increase very slowly with intensity, in agreement with the situation found in GX5-1 (ref. 3). The apparent lack of any correlation of full width at half maximum (FWHM) and intensity in the case of Cyg X-2 is not surprising as there the intensity was observed to change only by ~ 300 counts s^{-1} compared with $\sim 1,000$ counts s^{-1} for GX5-1.

The crude description of the corona presented here is inadequate, however, since it produces a period range for the QPOs that is too broad. Equations (4) and (5) also predict an unreasonably steep dependence of ν on θ unless the angle is sufficiently large that the $\theta \approx z/r$ assumption in equation (3) no longer holds. This may be rectified by allowing the illumination at the base of the corona, f , to vary with optical depth and thereby introducing a nonlinearity that localizes the $\tau=1$ surface in radius and thus in frequency. When the corona is optically thin, we would expect that the X-ray flux scattered onto the disk is proportional to the amount of scattering material above the disk. Rough calculations in which f varies linearly with τ indicate that an abrupt change in optical depth can result, for example, for $L=10^{38}$ erg s^{-1} , $\beta=8 \times 10^9$ and an illumination fraction f of 10^{-7} (at 10^6 cm) the optical depth increases dramatically at $r=4 \times 10^7$ cm. Detailed, self-consistent, radiative-hydrodynamical calculations are required to substantiate this.

We expect that such calculations would reproduce the overall behaviour of the inner corona outlined here.

QPOs have been observed for some time in cataclysmic variables¹⁷. It has been suggested that they are vertical oscillations of the (cold) accretion disk¹⁸. Scaling that model directly to a more compact object such as a neutron star or black hole leads to oscillation frequencies about an order of magnitude above those yet observed. We obtained lower frequencies because a significant corona exists only at relatively large radii where the radiation temperature corresponds to the local free-fall temperature.

Unlike previous models for QPOs involving a pulsar magnetosphere^{7,8}, our model neither requires nor precludes the existence of coherent pulsations at the rotation frequency of the neutron star. The Compton heated corona model does predict a complicated behaviour of the QPO phenomenon with spectral changes within the source. At a simplistic level one would not expect QPOs when the X-ray source is too soft (small corona) or when it is too hard (the case where there is a dense coronal wind¹⁹) or when a large magnetosphere from the central object extends out to the radius of the expected corona.

Finally, we note that the QPOs are expected to be easily observed in those systems with low inclinations, that is, seen at small angles from the symmetry axis of the disk. Eclipsing systems, which are those where there is a clear evidence for coronas, should not show QPOs because our line-of-sight is unlikely to penetrate to the inner disk regions.

We thank Douglas Gough and James Pringle for discussions. C.B.B. acknowledges the receipt of a SERC research fellowship. A.C.F. thanks the Royal Society for support.

Received 21 October; accepted 13 December 1985.

- White, N. E. *et al. Astrophys. J.* **247**, 994–1002 (1981).
- McClintock, J. E., London, R. A., Bond, H. E. & Grauer, A. D. *Astrophys. J.* **258**, 245–253 (1982).
- van der Klis, M. *et al. Nature* **316**, 225–230 (1985).
- Hasinger, G., Langmeier, A., Sztajno, M., Trumper, J., Lewin, W. H. G. & White, N. E. Preprint, MPE, Garching (1985).
- Middleditch, J. & Friedhorsky, W. *IAU Circ. No.* 4060 (1985).
- van der Klis, M. *On High Energy Phenomena Around Collapsed Stars, Cargèse* (NATO ASI, 1985).
- Alpar, M. A. & Shaham, J. *Nature* **316**, 239–241 (1985).
- Lamb, F. K., Shibasaki, N., Alpar, M. A. & Shaham, J. *Nature* **317**, 681–687 (1985).
- Motch, C., Ilovaisky, S. A. & Chevalier, C. *Astr. Astrophys.* **109**, 11–14 (1982).
- Motch, C., Ricketts, M. J., Page, C. G., Ilovaisky, S. A. & Chevalier, C. *Astr. Astrophys.* **119**, 171–176 (1983).
- Maehima, Y. *et al. Astrophys. J.* **285**, 712–720 (1984).
- Jones, C. & Forman, W. *Astrophys. J. Lett.* **209**, L131–L135 (1976).
- Fabian, A. C., Guilbert, P. W. & Ross, P. P. *Mon. Not. R. astr. Soc.* **199**, 1045–1051 (1982).
- White, N. E. & Holt, S. S. *Astrophys. J.* **257**, 318–337 (1982).
- Pringle, J. E. *A. Rev. Astr. Astrophys.* **19**, 137–162 (1981).
- Branduardi-Raymont, G., Kylafis, N. D., Lamb, D. Q. & Mason, K. O. *Astrophys. J.* **235**, 1153–1158 (1980).
- Robinson, E. L. & Nather, R. E. *Astrophys. J. Suppl.* **39**, 461–480 (1979).
- van Horne, H. M., Wesemael, F. & Winget, D. E. *Astrophys. J. Lett.* **235**, L143–L147 (1980).
- Begelman, M. C., McKee, C. F. & Shields, G. A. *Astrophys. J.* **271**, 70–88 (1983).

Pair production instabilities as a source of X-ray flares from accreting black holes

P. Moskalik & M. Sikora

N. Copernicus Astronomical Center, Polish Academy of Science, ul. Bartycka 18, 00-716 Warsaw, Poland

Several active galaxies emit most of their energy at $h\nu > 100$ keV (ref. 1). Observations suggest that their high-energy spectra are strongly variable and are cut off or broken at ~ 1 MeV (see for example, refs 2–6 and references therein). A possible explanation of these properties is that the high-energy radiation is produced in the central region (radius $R < 10R_s$, where R_s is the Schwarzschild radius) of the two-temperature plasma accreting onto a supermassive black hole^{7,8}. There, the e_+e_- pair creation can

be very efficient and, according to our calculations, at a certain luminosity range around $L_{\text{cr}} = 0.01 L_{\text{Edd}}$ (where L_{Edd} is the Eddington luminosity) it gives rise to a plasma instability. We show that for a proton temperature close to its virial value, $kT_p^{\text{virial}} \approx GMm_p/R$ (where m_p and M are masses of proton and of the central object respectively), this pair-production instability leads to cyclic variations of the accretion flow, during which high-energy flares are produced. We propose here that the large-amplitude luminosity changes noted in hard X rays and soft γ rays²⁻⁶, as well as the soft X-ray outbursts observed in several active galactic nuclei (AGNs)⁹⁻¹¹ result from the above mechanism. The same scenario may also be responsible for the short-timescale quasiperiodic variability reported in some proposed galactic black holes¹²⁻¹⁵.

Pair processes in thermal plasmas have been widely investigated during the past few years. The pair equilibrium states of the homogeneous static plasma cloud of finite radius R , in particular, have been studied (see ref. 16 for review), but without a specified electron heating mechanism. We assume that plasma is in a two-temperature state, with $T_p \gg T_e$ (the proton and electron/positron temperatures respectively), and that electrons/positrons cooled through comptonized thermal bremsstrahlung are reheated through Coulomb interactions with protons. We approximate the proton-to-electron energy transfer rate $\dot{u}_{\text{ep}}$, the electron cooling rate $\dot{u}_{\text{cbr}}$ and the pair creation rate $\dot{n}_+^{\text{cre}}$ using slightly modified formulas of Guilbert and Stepney¹⁷ (equations (11), (15) and (16)). For the annihilation rate $\dot{n}_+^{\text{ann}}$ we take the formula given by Svensson¹⁸ (equation (59)). Treating T_p as a model parameter, we obtain equilibrium solutions (satisfying the pair balance condition, as well as the electron/positron energy balance condition) which may be presented in two diagrams: luminosity versus electron temperature (Fig. 1) and particle densities versus electron temperature (Fig. 2). The equilibrium curves reproduce quite well those obtained by M. Sikora and M. Zbyszewska (in preparation) with more sophisticated treatment. The qualitative character of our curves is unaffected by the details of approximations used. At very low luminosities the pair production and comptonization are negligible, so we have $n_e = n_+ + n_- \approx n_p$ and $\dot{u}_{\text{cbr}} \approx \dot{u}_{\text{br}} = \text{bremsstrahlung emissivity}$. Thus the electron/positron heating and cooling rates depend on the plasma density through the same factor n_p^2 , and the electron energy balance condition, $\dot{u}_{\text{ep}} = \dot{u}_{\text{br}}$, gives $T_e = \text{constant}$ (T_p). At higher luminosities, but still lower than $L_{\text{cr}} = 0.01 L_{\text{Edd}}$, the comptonization becomes efficient while the pair production is still low. The energy balance equation takes the form $\dot{u}_{\text{ep}} = A(T_e \tau_T) \dot{u}_{\text{br}}$, where τ_T is the Thomson optical depth and A is the 'Compton amplification factor' having the following general properties: $A > 1$, $\partial A / \partial T_e > 0$ and $\partial A / \partial \tau_T > 0$. Taking into account that $\tau_T \propto n_e \approx n_p$ and that $\partial \dot{u}_{\text{br}} / \partial T_e > 0$ while $\partial \dot{u}_{\text{ep}} / \partial T_e < 0$, one can find that $dn_p/dT_e < 0$. At luminosities $L \geq L_{\text{cr}}$, the 'optical thickness' to pair production through the $\gamma\gamma \rightarrow e_+e_-$ process is > 1 (ref. 19). Now we have $n_e \gg n_p$, and function $L(T_e)$ and $n_e(T_e)$ flatten following the pair balance requirement, $\dot{n}_+^{\text{cre}} = \dot{n}_+^{\text{ann}}$, and dn_p/dT_e even changes its sign. When L is significantly larger than L_{cr} , the equilibrium electron temperatures are much lower than $m_e c^2$. The pair-production efficiency decreases because only photons from the exponential Wien tail are energetic enough to create pairs, and, therefore dn_p/dT_e becomes negative again and all functions, $L(T_e)$, $n_e(T_e)$ and $n_p(T_e)$, are steep.

Let us now consider pair density perturbations at fixed n_p . Close to the pair equilibrium state⁷, the electron heating and cooling timescales are significantly shorter than the creation and annihilation timescales. Thus, a pair density perturbation, δn_+ , will be accompanied by an electron temperature perturbation of opposite sign. In projection onto the n_p - T_e plane (Fig. 2), the $n_p = \text{constant}$ perturbation may be represented as a horizontal shift into the nonequilibrium region, where the term $\dot{n}_+^{\text{cre}} - \dot{n}_+^{\text{ann}}$ has the same sign as δn_+ if $(dn_p/dT_e)_{\text{eq}} > 0$, and an opposite sign if $(dn_p/dT_e)_{\text{eq}} < 0$. Thus, one can conclude that equilibrium

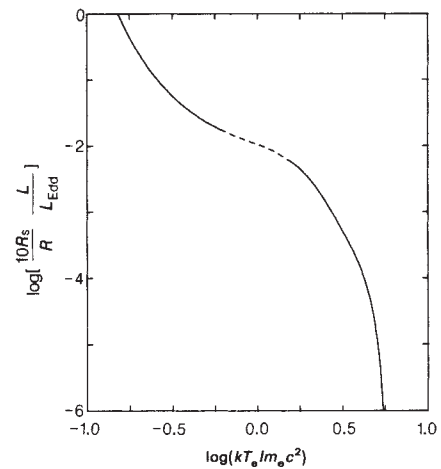


Fig. 1 The equilibrium luminosity of the two-temperature plasma cloud with $kT_p = 60 m_e c^2$ versus electron temperature. For $R = 10 R_s$ unstable solutions (indicated by the dashed line) appear at $L \sim 0.01 L_{\text{Edd}}$.

solutions corresponding to the range of parameters where $dn_p/dT_e > 0$ are unstable, whereas others are stable.

Hereafter we assume that our idealized plasma cloud represents the central part ($R \leq 10 R_s$) of matter accreting onto a black hole. We adopt the 'gravothermal' accretion scenario suggested by Meszaros and Ostriker²⁰ for spherically symmetrical flow and proposed⁷ also to apply in the case of accretion through a geometrically thick disk. In such a scenario, the two-temperature plasma, with a proton temperature close to its virial value, accretes on the timescale t^{acr} , corresponding to the energy transfer rate from protons to electrons and positrons. This timescale depends on n_e and/or T_e and, therefore, any variations in these parameters lead to variations of accretion velocity, as well as of the proton density. Thus, to study the time behaviour of the accretion flow undergoing pair instabilities we integrate the system of four differential equations describing: 1° pair creation and annihilation; 2° photon production, absorption and escape; 3° electron/positron heating and cooling; and 4° the proton density changes. The first three equations have been taken in approximate forms, similar to those used by Guilbert and Stepney¹⁷:

$$\frac{dn_+}{dt} = \dot{n}_+^{\text{cre}} - \dot{n}_+^{\text{ann}} \quad (1)$$

$$\frac{dn_j}{dt} = \frac{\dot{u}_{\text{cbr}}}{kT_e} - \frac{3n_j}{(1 + \tau_T)} \frac{c}{R} - 2 \frac{dn_+}{dt} \quad (2)$$

$$\frac{du_e}{dt} = \dot{u}_{\text{ep}} - \dot{u}_{\text{cbr}} \quad (3)$$

where $u_e \approx n_e kT$, $\tau_T = n_e R \sigma_T$ where σ_T is the Thomson cross-section, n_j is a photon density and n_+ is a positron density, equal to a pair density because of the neutrality condition. For the fourth equation we use

$$\frac{dn_p}{dt} = \lambda_0 - \frac{n_p}{t^{\text{acr}}} \quad (4)$$

where λ_0 is related to the rate of the matter inflow from the external regions ($R > 10 R_s$) and in our calculations is assumed to be constant. Linearizing the equations about an equilibrium state, we determine the range $\Delta \lambda_{0s}$ of λ_0 in which this equilibrium is unstable. This range is found to be only slightly narrower than that determined by the condition $dn_p/dT_e > 0$ (see dashed lines on Figs 1 and 2). This is so, because n_p changes on the accretion timescale, which is significantly longer than the pair-process timescales and our previous qualitative analysis is not drastically affected by $\delta n_p \neq 0$. Integrating the equations for

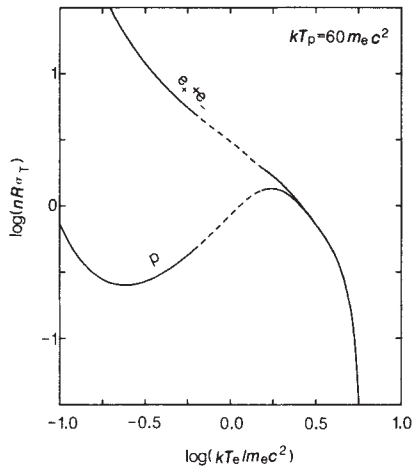


Fig. 2 The equilibrium electron/positron and proton densities versus electron temperature for $kT_p = 60m_e c^2$. Dashed lines indicate the unstable solutions.

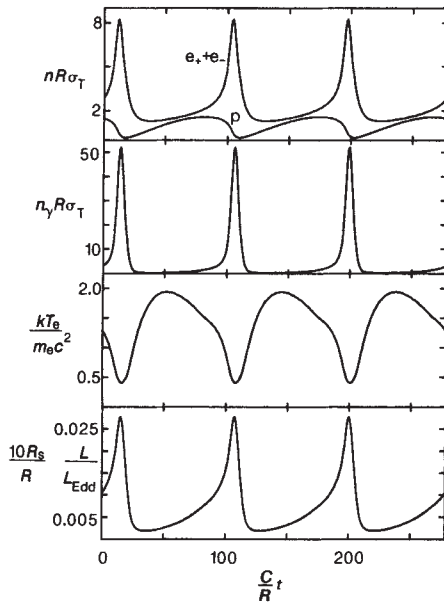


Fig. 3 The typical limit cycle ($L = 0.008 L_{\text{Edd}}$). The time variations of: a, the electron/positron and proton densities; b, the photon density; c, the electron temperature and d, the bolometric luminosity are plotted. The flare-like events occur during the pair-production runaway.

$\lambda_0 \in \Delta\lambda_{\text{us}}$, we see that all trajectories converge to a strictly periodic limit cycle. The exact form of this cyclic solution was found through the relaxation technique due to Stellingwerf²¹ using a numerical code described elsewhere²². We have found that for all values of $\lambda_0 \in \Delta\lambda_{\text{us}}$, such cycles are stable against small perturbations, but they are unstable or even do not exist for λ_0 beyond $\Delta\lambda_{\text{us}}$.

A typical solution is plotted in Fig. 3. Throughout most of the cycle protons accumulate because the electron/positron density is too low to dissipate effectively the protons' gravitational energy. This process continues until the photon density exceeds the critical value for pair-production instability. Now the electron/positron density grows rapidly, and as a consequence almost all the gravitational energy of the proton plasma stored so far is liberated in a short time interval and the high-energy ($h\nu > 100$ keV) flare is produced. The amount of energy emitted during the flare is roughly equal to:

$$E_{\text{fl}} \sim n_p \frac{GM_p}{R} \frac{4}{3} \pi R^3 \sim 10^{49} \tau_p \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right)^2 \left(\frac{R}{10 R_s} \right) \text{erg} \quad (5)$$

where $\tau_p = n_p R \sigma_T$ and M_{BH} is the mass of a black hole, while the timescale of the flare is found to be roughly consistent with the photon diffusion time, approximately equal to:

$$t_{\text{fl}} \sim \tau_T \left(\frac{R}{c} \right) \sim 10^4 \tau_T \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{R}{10 R_s} \right) \text{s} \quad (6)$$

Our calculations indicate that the recurrence period P is of the order of:

$$P \sim 100 \left(\frac{R}{c} \right) \sim 10^6 \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{R}{10 R_s} \right) \text{s} \quad (7)$$

which is a few times longer than the equilibrium accretion timescale. Note however, that the cycle period depends on λ_0 (roughly as $\sim 1/\lambda_0$) and can change by a factor of a few when modulated by a variable matter inflow.

Some part of the flare energy (equation (5)) is supposed to be transferred into softer spectrum bands through various effects not included in our crude model. For example, this can take place in the disk funnel that is filled during the flare with rapidly cooling pair plasma. The most dramatic luminosity changes are expected, however, for $h\nu > 100$ keV. Sporadic measurements of active galaxies show high-amplitude changes in their soft γ -ray flux (ref. 3 and references therein), but the available data are not sufficient to follow continuous time variations. More useful observations have been made in the medium and soft X-ray bands and flare-like events have been noted in several AGNs⁹. Repeated flares have been observed in NGC4151 (ref. 10) and in Centaurus A (ref. 11) with recurrence timescales of weeks, which are consistent with our estimations (equation (7)), if $M_{\text{BH}} \sim 10^8$ – $10^9 M_{\odot}$ is assumed. Changes in the accretion rate can lead to flaring irregularities and may even switch off this type of variability. Such an off/on behaviour probably takes place in Cen A (refs 11, 23). We believe that future observations will show that flares, sporadically noted in some other objects, are not exceptional, but are manifestation of continuous flaring activity. This supposition is supported by the fact that the flare timescales in these objects, as well as in NGC4151 and Cen A, are of the order of days and agree with the timescales indicated by equation (6) for the same range of black hole masses as before.

Of course, the range of AGN black hole masses is expected to be much larger ($10^6 M_{\odot}$ – $10^{10} M_{\odot}$) and, due to the correlation $t_{\text{fl}} \propto M_{\text{BH}}$, we should also observe flares on timescales much shorter than 1 day. The absence of such flares (with one or two possible exceptions)²⁴ can be explained by noting that, for objects with measured γ -ray luminosity L_{γ} , the ratio L_{γ}/L_x (where L_x is X-ray luminosity) is of the order of 100 (see ref. 25). Since the instability develops only in the narrow luminosity range around $L_{\text{cr}} = 0.01 L_{\text{Edd}}$ (see Fig. 1), the only AGNs expected to produce flares are those with L_{γ} close to L_{cr} . Thus, for flaring objects L_x should correlate with t_{fl} :

$$L_x \sim 0.01 L_{\gamma} \sim 10^{-4} L_{\text{Edd}} \sim 10^{42} \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \frac{\text{erg}}{\text{s}} \\ \sim \frac{10^{43}}{\tau_T} \left(\frac{t_{\text{fl}}}{10^5 \text{s}} \right) \left(\frac{10 R_s}{R} \right) \text{erg s}^{-1} \quad (8)$$

Hence, for $t_{\text{fl}} < 10^4$ s we obtain $L_x < 10^{42} \text{ erg s}^{-1}$, that is below the present detector sensitivities for typical distances to active galaxies. Equation (8) also explains why the high X-ray luminosity objects do not vary on timescales of days²⁶.

Since the solutions presented in Fig. 3 do not depend on the mass of the black hole, one may expect that the same mechanism also acts in the case of stellar black holes. Clearly, the corresponding timescales must be much shorter, for example, $t_{\text{fl}} \sim 10$ ms and $P \sim 0.1$ s for $M_{\text{BH}} \sim 10 M_{\odot}$. There is some observational evidence that galactic black-hole candidates Cyg X-1, Cir X-1 and GX339-4 exhibit sporadic quasi-periodic flaring activity^{12–15}. The estimated bolometric luminosities of these

objects (see Table 1 in ref. 12) seem to fall into the range appropriate for the pair instability. The shortest timescale of the variability (t_{fi}) agrees roughly with our predictions, but the recurrence period P does not agree, being of the order of ~ 1 s. We feel that this discrepancy does not rule out the possibility that the same instability mechanism works in such objects. Our quantitative model is very simplified, in particular, that it assumes spatial homogeneity.

Finally, it should be stressed the gravothermal scenario breaks down at luminosities higher than $0.1 L_{\text{Edd}}$ because the resulting accretion velocity becomes comparable to or even higher than the free-fall velocity. Then, a quite different scenario (such as accretion through an optically thick disk with a hot corona) must be considered, which may be related to the 'high states' occasionally noted in the galactic X-ray sources mentioned above and supposed to be permanent in BL Lacertae objects²⁷.

We thank Professor W. Dziembowski for comments, also Professor J. Smak for helping to improve the text. The computations reported here were made with PDP 11/45 computer donated to the N. Copernicus Astronomical Center by the US National Academy of Science, following the action of Dr C. R. O'Dell and Mr A. M. Baer.

Received 9 August; accepted 27 November 1985.

1. Rothschild, R. E. *et al. Astrophys. J.* **269**, 423–437 (1983).
2. Baity, W. A. *et al. Astrophys. J.* **279**, 555–562 (1984).
3. Bassani, L. & Dean, A. J. *Space Sci. Rev.* **35**, 367–398 (1983).
4. Bezler, M. *Astr. Astrophys.* **136**, 351–354 (1984).
5. Gehrels, N. *et al. Astrophys. J.* **278**, 112–118 (1984).
6. Ubertini, P., Bazzano, A., La Padula, C. & Polcavo, V. F. *Astrophys. J.* **284**, 54–59 (1984).
7. Sikora, M. & Zbyszewska, M. *Mon. Not. R. astr. Soc.* **212**, 553–564 (1985).
8. Takahara, F. & Kusunose, M. *Prog. theor. Phys.* (submitted).
9. Marshall, N., Warwick, R. S. & Pounds, K. A. *Mon. Not. R. astr. Soc.* **194**, 987–1002 (1981).
10. Lawrence, A. *Mon. Not. R. astr. Soc.* **192**, 83–94 (1980).
11. Lawrence, A., Pye, J. P. & Elvis, M. *Mon. Not. R. astr. Soc.* **181**, 93p–99p (1977).
12. Samimi, J. *et al. Nature* **278**, 434–436 (1979).
13. Margon, B., Lampton, M., Bowyers, S. & Cruddace, R. *Astrophys. J. Lett.* **169**, L23–L25 (1971).
14. Sadeh, D. *et al. Nature* **278**, 436–438 (1979).
15. Oda, M. *Space Sci. Rev.* **20**, 757–813 (1977).
16. Svensson, R. *Proc. Conf. X-ray and UV Emission from Active Galactic Nuclei*, (eds Brinkman, W. & Trumpler, J.) 152–163 (MPE Rep. 184 1984).
17. Guilbert, P. W. & Susan Stepney *Mon. Not. R. astr. Soc.* **212**, 523–544 (1985).
18. Svensson, R. *Astrophys. J.* **258**, 335–348 (1982).
19. Lightman, A. P., Giacconi, R. & Tananbaum, H. *Astrophys. J.* **224**, 375–380 (1978).
20. Meszaros, P. & Ostriker, J. P. *Astrophys. J. Lett.* **273**, L59–L63 (1983).
21. Stellingwerf, R. F. *Astrophys. J.* **192**, 139–144 (1974).
22. Moskalik, P. *Acta astr.* (in the press).
23. Terrel, J. *Proc. Conf. X-ray and UV Emission from Active Galactic Nuclei* (eds Brinkman, W. & Trumpler, J.) 102–104 (MPE Rep. 184 1984).
24. Allyn Tennant, F. & Mushotzky, R. F. *Astrophys. J.* **264**, 92–104 (1983).
25. Bassani, L. & Dean, A. J. *Astrophys. Space Sci.* **100**, 457–460 (1984).
26. Zamorani, G., Giommi, P., Maccaacaro, T. & Tananbaum, H. *Astrophys. J.* **278**, 28–36 (1984).
27. White, N. E., Fabian, A. C. & Mushotzky, R. F. *Astr. Astrophys.* **133**, L9–L11 (1984).

Measuring interstellar dust grains from the haloes of binary X-ray sources

Yueming Xu*, Richard McCray* & Richard Kelley†

* Joint Institute for Laboratory Astrophysics, University of Colorado, Boulder, Colorado 80309, USA

† Department of Physics and Astronomy, University of Maryland, and Laboratory for High Energy Astrophysics, NASA-Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Overbeck¹ first predicted that the images of cosmic X-ray sources might have observable haloes due to coherent forward scattering by interstellar dust grains, and it was recognized that the observations of these haloes would provide a powerful tool for analysing the size distribution and composition of the grains^{2–5}. Observations with the Einstein observatory have provided evidence of faint haloes extending to a few arc minutes beyond the X-ray images of supernova remnants^{6,7} and several compact galactic sources^{8–11}. The evidence that the haloes are actually due to interstellar grain

scattering appears substantial, but it is subject to some uncertainty resulting from the dependence of the point response function (PRF) of the telescope on the source spectra (the observed haloes are brighter than the predicted PRF by factors of approximately 1.5–2). In this letter we show how observations of the time-dependence of the halo X-ray emission may be analysed to determine the location, size distribution, and composition of interstellar grains. Indeed, it should be possible to observe and analyse the halo emission during eclipse even without imaging the source, and we suggest that such haloes probably account for the soft X-rays that have been observed during eclipses or dips in several binary systems.

Here we quote results for a simple but plausible specific model; detailed derivations of these and more general results will be presented elsewhere. The model consists of a variable compact X-ray source at a distance D from the observer. At a distance r from the observer is a dust cloud of thickness L or several clouds distributed over a range of distance L . The optical depth for scattering of X-rays by the grains is assumed to be less than one, so that multiple scattering may be neglected. The dust grains are assumed to be spheres composed of silicate (MgSiO_3) or graphite with a number density size distribution $dn_g/da = Ka^{-m}$ for $a_{\text{min}} < a < a_{\text{max}}$, and zero otherwise. The optical extinction is dominated by grains with radius $a \approx 0.1 \mu\text{m}$, but the infrared to ultraviolet wavelength dependence of the extinction indicates that $m = 3.5$ with a very uncertain value of $a_{\text{max}} \approx 0.3\text{--}1 \mu\text{m}$ (ref. 12). The X-ray differential scattering cross-section of the grains (absorption in the grain is fairly unimportant) is represented by the Rayleigh-Gans approximation¹³, so that a photon of wavelength λ may be scattered by a typical angle $\theta_1 \approx \lambda/\pi a$ by a grain of radius a . Such photons (with energy E_{keV}) will appear in a halo around the point source with angular size

$$\alpha_1 \approx 1.4 \text{ arc min } E_{\text{keV}}^{-1} (a/1 \mu\text{m})^{-1} (1-r/D) \quad (1)$$

Scattered photons arrive later than unscattered photons with a time delay given by

$$t \approx 1.2 \text{ h } (\alpha/1 \text{ arc min})^2 (r/1 \text{ kpc}) (1-r/D)^{-1} \quad (2)$$

Thus, during a total eclipse, the point source (and, of course, its PRF) disappears but the halo persists for several hours, as a dark circle grows within its image according to equation (2). Note that the distance of the dust cloud can be inferred directly from observations of the growth of this dark circle if the distance of the source is known^{14,15}.

Rapid fluctuations or pulsations in the brightness of the compact source are washed out in the halo image by the distribution of time delays in the dust cloud. The minimum fluctuation period that can be observed in the halo is given by

$$P_{\text{min}} \approx 4.3 \text{ s } (\alpha/1 \text{ arc min})^2 (L/1 \text{ pc}) (1-r/D)^{-2} \quad (3)$$

Thus, even without a total eclipse, a halo caused by dust grain scattering may be distinguished from the PRF by its failure to track the rapid variability of the central source¹¹.

Assuming that the source is fairly steady before eclipse, the ratio of the integrated brightness, $B_h(E)$, of the halo at a given photon energy to that, $B_s(E)$, of the unscattered source before eclipse is given by $[\exp(\tau_X) - 1]$, where τ_X , the total optical depth to X-ray scattering, is given by (for $m < 5$)

$$\tau_X(E) \approx 1.0 A_V (a_{\text{max}}/1 \mu\text{m})^{5-m} E_{\text{keV}}^{-2} \quad (4)$$

where $A_V \approx 3 E_{\text{B-V}}$ is the optical extinction and we have assumed silicate grains with optical extinction efficiencies given by Draine and Lee¹⁶. (The numerical coefficient of equation (4) is uncertain by a factor of ~ 2 ; it is smaller by a factor of ~ 4 if the grains are composed of graphite.) Thus, even without an imaging X-ray telescope, a halo caused by interstellar grain scattering may be recognized readily by its soft spectrum immediately after eclipse, which is steeper than that of the central source by E^{-2} . Note that the halo is dominated by scattering from the largest grains.

Note also that both the central image and the halo will suffer approximately the same interstellar photoelectric absorption if the dust clouds are fairly uniform over a lateral extent $\sim \alpha_1(E)$. The latter assumption can be checked by measuring the azimuthal uniformity of the halo.

The Rayleigh-Gans approximation will be accurate provided that the quantity $\delta \approx 6.7 (a_{\max}/1 \mu\text{m})/E_{\text{keV}} < 1$, and will overestimate the forward scattering cross-section for greater values of δ (ref. 17). Our calculations using the more accurate anomalous dispersion approximation¹³ show that the resulting error in equation (4) is less than a factor of 2 for $\delta < 3$.

As the eclipse progresses and the central halo image darkens according to equation (2), the high-energy X-rays that are scattered to small angles by large grains vanish first; this causes the integrated halo spectrum to develop a break at a photon energy

$$E_1(t) \approx 1.5 \text{ keV} (a_{\max}/1 \mu\text{m})^{-1} (t/1 \text{ h})^{-1/2} \times [(r/1 \text{ kpc})(1-r/D)]^{1/2} \quad (5)$$

above which $\tau_X(E) \propto E^{m-7} t^{(m-5)/2}$ for $3 < m < 5$ and $\tau_X(E) \propto E^{-4} t^{-1}$ for $m < 3$. Thus, the size distribution of the grains can be inferred from the time dependence of the spectrum.

Alternatively, with an imaging telescope the grain size distribution may be inferred directly from the halo brightness profile at a given photon energy¹⁰. For $\alpha < \alpha_1(E, a_{\max})$ [cf. eq. (1)] the brightness is fairly uniform and for $\alpha > \alpha_1$ and $m > 3$ it is given by

$$I(E, \alpha) \propto \alpha^{m-7} \quad (6)$$

The residual signal from the halo after eclipse will probably be dominated by the softest photons that are detected. The energy of such photons is determined either by the detector response or by the X-ray opacity of the interstellar medium. If the latter factor prevails, the residual signal is nearly independent of the column density of interstellar matter, because as A_V decreases, so does the soft X-ray cutoff, $E_{\min}$, due to interstellar absorption¹⁸. Assuming standard values for cosmic abundances and for the gas-to-grain ratio¹⁹, we may estimate that the signal is dominated by photons with $E = E_{\min} \approx 0.7 \text{ keV } A_V^{0.4}$, for which equation (4) becomes

$$\tau_X(E_{\min}) \approx 2A_V^{0.2} (a_{\max}/1 \mu\text{m})^{5-m} \quad (7)$$

Thus, multiple scattering (which we have neglected here) may become important for $E \approx E_{\min}$ if $a_{\max} > 0.4 \mu\text{m}$.

It may be possible to infer the composition of the grains from the spectrum of the halo by observing spectral features in $\tau_X(E)$ near the K-shell photoionization thresholds of the elements comprising the grains^{4,5,20}. The haloes will become dim just below ionization threshold, then sharply increase in fractional brightness, $\tau_X(E)$, as the photon energy increases past the absorption edge of the relevant atom. For example, for $a_{\max} = 0.3 \mu\text{m}$ we calculate (using the anomalous dispersion approximation) that $\tau_X(E)$ will increase by a factor of ~ 6 across the 0.29-keV carbon K edge if the grains are graphite and by a factor of ~ 1.4 across the 0.53 keV oxygen K edge if the grains are composed of silicates.

The effects of X-ray scattering by interstellar dust may already be evident in several observations of X-ray binaries during X-ray eclipses or dips. Schreier *et al.*²¹ were the first to report a residual flux from Cen X-3 during eclipse, but they were unable to determine the spectral or temporal behaviour. More recently, White²² has presented eclipsed and uneclipsed spectra of Cen X-3 using data from the Einstein Solid State Spectrometer (SSS) and OSO-8. The SSS photon spectrum appears to have a slope of $\sim E^{-2}$ during eclipse compared to a fairly flat uneclipsed spectrum, consistent with equation (4). The Einstein Monitor Proportional Counter shows a residual flux of $\sim 20\%$ of the uneclipsed flux at a characteristic photon energy of $\sim 3 \text{ keV}$. This is consistent with equation (4) for $A_V = 4.4$ if $a_{\max} \sim 0.5 \mu\text{m}$. According to equation (5), the energy break at mid-X-ray eclipse

should occur at $\sim 2 \text{ keV}$; this should be readily detectable with EXOSAT and future X-ray missions.

Similarly, the intensity dips of Cyg X-1, which are due to photoelectric absorption by gas in the binary system, also show a residual soft spectral component with a 2 keV flux of approximately 15% of the unobscured flux²³. This result agrees very well with that¹¹ obtained from analysis of the Einstein High Resolution Imager image of Cyg X-1 and is consistent with equation (4) for $A_V = 3.3$ and $a_{\max} \approx 0.3 \mu\text{m}$.

Residual soft X-ray emission has also been observed during dips of the sources 4U1915-05 (ref. 24) and X1755-338 (ref. 25). This emission has been interpreted to imply either partial covering of the source or very low metallicity of the obscuring gas. We suggest alternatively that this emission might be due to persistent dust haloes, as seems to be the case for Cyg X-1.

With $A_V = 19$, Cyg X-3 should have a very bright X-ray halo. For example, for silicate grains with $a_{\max} = 0.3 \mu\text{m}$, equation (7) predicts $\tau_X(E_{\min}) \approx 0.6$ at the soft X-ray cutoff energy $E_{\min} \approx 2.3 \text{ keV}$. This is substantially greater than the value $\tau_X \approx 0.30$ that has been inferred from observations with the Einstein Imaging Proportional Counter¹⁰, but it may be consistent with the soft component in the 2-4 keV spectrum that is not modulated with the 4.8 h period²⁶.

Several other binary sources, including 4U0900-40, 4U1700-37, and 4U1538-52, have been observed during X-ray eclipses with OSO-8, HEAO-1, and the Einstein Observatory. It is unlikely that the residual soft X-ray flux during eclipse of these sources could be due to electron scattering in the stellar wind of the companion star. The spectrum resulting from that process should resemble that of the uneclipsed source with a soft X-ray cutoff due to photoelectric absorption by ions in the wind (T. Kallman & J. Swank, in preparation).

It should be clear that astronomers cannot safely ignore the effects of interstellar grain scattering in interpreting the soft X-ray emission from these and other variable sources. More important, it seems likely that further analysis of these data and future spectroscopic and imaging data from the EXOSAT, ROSAT, BBXRT, and AXAF observatories may provide definitive information on the composition and size distribution of interstellar grains.

We thank John Black, Michael Bode, Bruce Draine, Paul Gorenstein, Tim Kallman, Jean Swank and Nick White for advice and discussions. This research was supported in part by a grant from NASA.

Received 27 August; accepted 20 November 1985.

- Overbeck, J. W. *Astrophys. J.* **141**, 864-866 (1965).
- Slysh, V. I. *Nature* **224**, 159-161 (1969).
- Hayakawa, S. *Prog. theor. Phys.* **43**, 1224-1230 (1970).
- Martin, P. G. & Sciama, D. W. *Astrophys. Lett.* **5**, 193-196 (1970).
- Martin, P. G. *Mon. Not. R. Astr. Soc.* **149**, 221-235 (1970).
- Stewart, G. C., Fabian, A. C. & Seward, F. D. in *Supernova Remnants and their X-Ray Emission* (eds Danziger, J. & Gorenstein, P.) 59-64 (Reidel, Dordrecht, 1983).
- Mauche, C. W. & Gorenstein, P. in *Proc. Workshop on the Crab Nebula and Related Supernova Remnants* (eds Kafatos, M. C. & Henry, R. B. C.) 81-88 (Cambridge University Press, 1985).
- Rolf, D. P. *Nature* **302**, 46-48 (1983).
- Catura, R. C. *Astrophys. J.* **275**, 645-651 (1983).
- Spitzer, L. Jr & Gorenstein, P. *Astrophys. J.* (in the press).
- Bode, M. F., Friedhorsky, W. C., Norwell, G. A. & Evans, A. *Astrophys. J.* **299**, 845-851 (1985).
- Mathis, J. S., Rump, W. & Nordsieck, K. H. *Astrophys. J.* **217**, 425-433 (1977).
- van de Hulst, H. C. *Light Scattering by Small Particles* (Wiley, New York, 1957).
- Trümper, J. & Schönfelder, V. *Astron. Astrophys.* **25**, 445-450 (1973).
- Spiegelhauer, H. & Trümper, J. *J. Br. Interplanetary Soc.* **28**, 319-325 (1975).
- Draine, B. T. & Lee, H. M. *Astrophys. J.* **285**, 89-108 (1984).
- Alcock, C. & Hatchett, S. *Astrophys. J.* **222**, 456-470 (1978).
- Morrison, R. & McCammon, D. *Astrophys. J.* **270**, 119-122 (1983).
- Spitzer, L. Jr *Physical Processes in the Interstellar Medium* (Wiley: New York, 1978).
- Henke, B. L., Lee, P., Tanaka, T. J., Shimabukuro, R. L. & Fujikawa, B. K. *Atomic Data and Nuclear Data Tables* **27**, 1-144 (1982).
- Schreier, E., Levinson, R., Gursky, H., Kellogg, E., Tananbaum, H. & Giacconi, R. *Astrophys. J. Lett.* **172**, L79-89 (1972).
- White, N. E. in *Interacting Binary Stars*, *Proc. NATO Institute for Advanced Studies* (eds Eggleton, P. & Pringle, J.) 249-287 (Reidel, Dordrecht, 1985).
- Kitamoto, S. *et al. Publ. Astr. Soc. Japan* **36**, 731-740 (1984).
- White, N. E. & Swank, J. H. *Astrophys. J.* **253**, L61-L66 (1982).
- White, N. E. *et al. Astrophys. J.* **283**, L9-L12 (1984).
- White, N. E. & Holt, S. S. *Astrophys. J.* **257**, 318-337 (1983).

Time-integrated energy budget of a solar activity complex

G. A. Chapman, A. D. Herzog & J. K. Lawrence

San Fernando Observatory, Department of Physics and Astronomy,
California State University, Northridge, California 91330, USA

The 0.1–0.3% dips in solar irradiance during disk passages of large sunspot groups suggest the possibility of fluctuations in the solar luminosity. This raises the question of whether the energy not radiated by the dark sunspots is stored within the Sun for long periods of time, or is radiated by faculae during the several-month lifetime of a solar activity complex^{1–3}. Here we examine the sunspot and facular contributions to luminosity fluctuations due to a solar activity complex over its lifetime from June to November 1982. Both direct, photometric observations of irradiance fluctuations and modelled 'proxy' fluctuations based on published sunspot and calcium plage areas are used. We find that the total facular energy excess is between 70 and 120% of the sunspot deficit of $\sim 10^{37}$ erg. Thus, at a minimum, a major portion of the missing sunspot flux is radiated by faculae, and energy balance or even an excess is possible. This work differs from earlier studies^{2–9} in that our data cover a longer period of time, more photometric data are included, and our analysis considers the effect of partial occultation of active regions by the solar limb.

The activity complex examined here appeared in June 1982 as Big Bear Solar Observatory (BBSO) active region number 18422. It reappeared in July as 18474, in August as 18511, in September as 18550, 551 (NOAA 3885, 3886), in October as 18587, 592, and twice in November as undesignated, scattered plage. No substantial sunspots clearly associated with the plage were reported from October onwards.

Our direct photometric data consist of two-dimensional photoelectric images of the activity complex, made with the San Fernando Observatory (SFO) 512-element Reticon diode array, vacuum spectroheliograph and 28-cm vacuum telescope. These were made with a pixel spacing of 0.94 arc s and a bandwidth of 1.5 Å in clean continuum at 6,264 Å. After limb darkening and other corrections are numerically applied, irradiance fluctuations may be directly calculated by summing the intensities of sunspot and facular pixels. Details are provided in refs 4, 5. Good coverage of our activity complex was obtained, with several images per day, during disk passages in July, August and September 1982.

Good photometric data are not available for the remaining disk passages, so modelled irradiance variations based on measured plage and sunspot areas are substituted. The models are

$$\text{PSI} = -C_s A_s \mu (3\mu + 2) \quad (1)$$

and

$$\text{PFI} = C_p A_p (\mu - 3\mu^2 + 2) \quad (2)$$

where PSI and PFI are the sunspot and facular irradiance variations⁶ in parts per 10^6 (p.p.m.) calibrated for wavelength 5,250 Å; C_s and C_p are empirically determined constants, and A_s and A_p are the region's sunspot and calcium plage areas, respectively; μ is the cosine of the region's heliocentric angle. With the exceptions noted below, sunspot and plage areas used are those published in the World Data Center A Solar-Geophysical Data Bulletin. We have adopted $C_s = 0.164$ as estimated elsewhere^{1,10}. C_p is smaller and harder to determine⁹. Early values of $C_p = 0.01$ are too small. Using the SFO extreme limb photometer (ELP), Chapman and Meyer¹⁰ have found, for 5,250 Å, $C_p = 0.0185$ with a 1σ error of 15%, and we adopt this value. The proxy facular emission is proportional to this number, so our final results are easily adjusted to fit other values of C_p .

By integrating PSI and PFI over $0 < \mu < 1$ we obtain instantaneous luminosity fluctuations due to this activity complex in p.p.m. of the solar luminosity^{4,5}

$$\Delta L_s = -C_s A_s \quad (3)$$

and

$$\Delta L_f = \frac{3}{4} C_p A_p \quad (4)$$

Examination of published areas of active-region plage and sunspots shows a decrease near the limb (for $\mu \leq 0.5$); we interpret this as a loss of visible area, as a region extending over $\sim 30^\circ$ of longitude is partially occulted by the limb. Since the luminosity fluctuations are independent of the Earth's location, which causes the occultation, we must use the full area in equations (3) and (4). The photometric luminosity fluctuations of refs 4, 5 are based on direct irradiance measurements and have not been corrected for the occultation effect. In the case of faculae, the earlier photometric luminosity fluctuation should be increased by 30%. Since sunspot contributions vanish at the limb where this effect is largest, the correction for sunspots is only a 6% increase in magnitude.

Our photometric data are monochromatic at 6,264 Å, and our proxy irradiance fluctuations are calibrated to match monochromatic ELP data at 5,250 Å. We estimate the bolometric luminosity fluctuations from their values at these particular wavelengths by assuming that faculae, sunspots and quiet photosphere have black-body spectra at temperatures of 6,200, 4,000 and 5,800 K, respectively. This gives fluctuation magnitude corrections for photometric faculae of +6%, photometric sunspots –7%, proxy faculae –15%, and proxy sunspots –12%. We have also examined the effect of atmospheric and instrumental stray light on our photometric facular and sunspot contrasts, and we have estimated that stray light decreases the observed magnitudes of irradiance fluctuations of both faculae and sunspots by about 15% (ref. 4).

Satellite observations show that significant chromospheric emission, beyond that estimated by our bolometric corrections, is associated with Ca plage. Combination of the ultraviolet plage contrasts published by Lean¹¹ with the ultraviolet quiet Sun irradiance spectrum of Samain¹² indicates that an ultraviolet luminosity correction $\Delta L_u = C_u A_p$ with $C_u = 0.0015$ should be added to ΔL_f .

Figure 1 shows the corrected facular and sunspot luminosity excesses and deficits for our activity complex as a function of

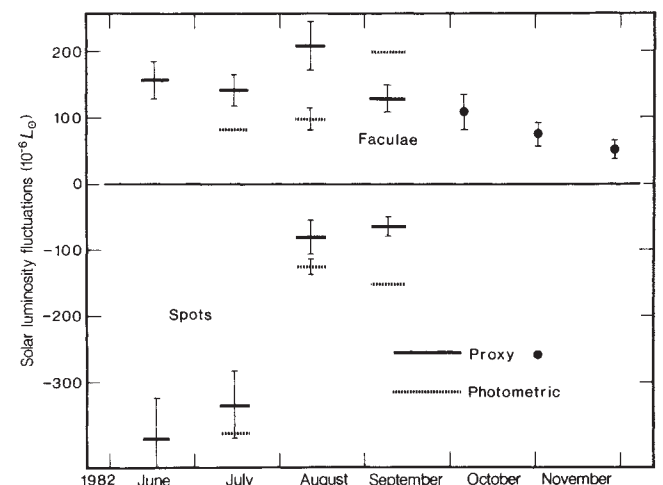


Fig. 1 Facular and sunspot luminosity fluctuations by 1982 date. Each symbol represents the integration, over angle, of photometric or modelled (proxy) irradiance fluctuations of a sequence of active regions making up an 'activity complex'. The width of the bars represents the time over which the region could be observed. The error bars include random errors and solar changes. The facular energy excess is roughly equal to the energy deficit of spots.

time. The horizontal bars show averaged luminosity fluctuation during disk passages of the complex, both for proxy data based on Solar Geophysical Data Bulletin (SGD) areas with $C_p = 0.0185$ and for photometric data when available. The single points are proxy values based on single-day Ca plage areas obtained directly from BBSO. The error bars represent the estimated random errors. The vertical error bars on the proxy facular luminosity variations represent the 15% 1σ uncertainty in C_p mentioned above, together with the day-to-day fluctuation in A_p . The error bars on the proxy sunspot luminosity variations represent a corresponding 5% uncertainty in C_s and the day-to-day fluctuation in A_s . The error bars on the August photometric luminosity fluctuations represent the formal 1σ errors estimated in ref. 4 and are probably typical of the July and September results. Systematic errors appear to be more important than random errors; the photometric data, for example, give greater sunspot deficits than do the proxy data.

To compute the total energy deficits and excesses we must integrate the luminosity fluctuations over time. Because of their equal spacing in time the simplest estimate of the integrals is obtained by assuming that each luminosity determination holds for one rotation of the activity complex, and then summing. This integration technique is equivalent to assuming that the complex was born at midpoint on the back of the Sun with the sunspot area observed in June. Since sunspot groups typically grow to maximum area over periods of ~ 10 days (ref. 13), this method may slightly overemphasize the sunspot deficit. The plage was last observed as part of a detectable complex in November, but must continue to contribute as scattered plage to an energy excess. By fitting an exponential curve (with decay constant ~ 2 months) to the last three data points, we estimate this to be a 10% increment in the total facular energy excess.

When we now sum the energies using proxy data alone, we find that the facular energy excess is 110% of the sunspot deficit of $\sim 8 \times 10^{36}$ erg. When we use photometric data instead of proxy, when possible, we find an excess of 80% of the sunspot deficit of $\sim 9 \times 10^{36}$ erg. The expected random error in these percentages is $\sim 10\%$ so we estimate that facular emission amounts to 70–120% of the sunspot energy deficit of $\sim 10^{37}$ erg over the activity complex lifetime.

When we study the evolution of a particular activity complex, we can only assume that no unusual behaviour occurs when the complex is out of sight on the back of the Sun. One can remove this difficulty, however, by examining many regions on the front of the Sun. A sum of all published areas of the 636 active regions crossing the solar central meridian during June–September of 1977, 1980–81 and June–August 1982 give total areas $A_p = 6.6 \times 10^5$ p.p.m. and $A_s = 5.9 \times 10^4$ p.p.m. Use of our equations (3) and (4) and all of the proxy data corrections discussed above indicates that the overall facular energy excess is $110 \pm 20\%$ of the corresponding sunspot deficit, in agreement with our previous result. This result does not include systematic errors in A_p and A_s , or the uncertainty in the mathematical form of equations (1) and (2).

More work is needed to determine the exact energy budget within an activity complex. Such a determination is important for understanding the nature of sunspots and the solar dynamo. Our work shows, on the basis of photometric and modelled data, that energy balance is a realistic possibility. Improvement of this result will require the acquisition and analysis of a considerably larger body of ground-based and satellite data.

9. Schatten, K. H. *et al. Astrophys. J.* **294**, 689–696 (1985).
10. Chapman, G. A. & Meyer, A. D. *Sol. Phys.* (in the press).
11. Lean, J. L. *J. geophys. Res.* **89**, 1–9 (1984).
12. Samain, D. *Astr. Astrophys.* **74**, 225–228 (1979).
13. Bray, R. J. & Loughhead, R. E. *Sunspots*, 226 (Chapman and Hall, London, 1964).

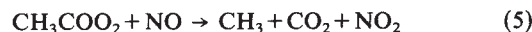
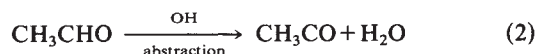
The spring maximum in photo-oxidants in the Northern Hemisphere troposphere

S. A. Penkett* & K. A. Brice*

Environmental and Medical Sciences Division,
Atomic Energy Research Establishment, Harwell, OX11 0RA, UK

Peroxy acetyl nitrate (PAN) is an excellent tracer of photochemical activity in the troposphere. We have obtained results at Harwell which show a large increase in the concentration in background air in springtime. We suggest that this indicates that tropospheric photochemistry may contribute to the spring maximum in the tropospheric ozone concentration in the Northern Hemisphere, which has traditionally been ascribed to increased transfer of air from the stratosphere.

PAN is a characteristic product of atmospheric photochemistry¹. It is formed when hydrocarbons other than methane are oxidized by hydroxyl radicals in the presence of molecular oxygen and nitrogen dioxide². In the case of propene (C_3H_6),



Propene is a characteristic hydrocarbon that has both natural and anthropogenic sources³. Most other hydrocarbons, including alkanes and aromatics, also give rise to PAN production during atmospheric oxidation⁴. The hydroxyl radicals in the clean troposphere are mainly produced from the photolysis of ozone in the small window of the light that penetrates the stratospheric ozone layer and overlaps the ozone absorption in the bands between 290 and 310 nm (ref. 5).

PAN is in a dynamic equilibrium with NO_2 and the CH_3COO_2 radicals (reaction (4)); it is removed from the atmosphere by reaction (5)⁶, by reaction with hydroxyl radicals⁷ and by absorption on the ground surface⁸. Its solubility in water is low so that it is not washed out by rain or removed to any extent by absorption into the ocean⁹.

Many measurements of PAN have been made in polluted atmospheres^{10–12}. Measurements in clean-air regimes are much more scarce. The first were reported in 1974 over the Atlantic Ocean¹³. Recently, measurements have been made in the free troposphere over the Pacific¹⁴ and over the European continent¹⁵. They all show the presence of substantial amounts of PAN [> 100 parts per 10^{12} by volume (p.p.t.v)] in background Northern Hemisphere air.

We now report some clean-air measurements made at ground level at Harwell ($51^\circ 37' N$, $1^\circ 18' W$) in rural southern England. The site has been well characterized by a series of measurements of halocarbons made between 1974 and 1981, when it was shown that the site was exposed to clean background air, typical of the Northern Hemisphere troposphere, for $\sim 10\%$ of the time¹⁶.

* Present addresses: School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK (S.A.P.); Concord Scientific Corporation, Tippet Road, Toronto, Ontario, Canada (K.A.B.).

Received 14 August; accepted 5 December 1985.

1. Willson, R. C. *et al. Science* **211**, 700–702 (1981).
2. Eddy, J. A., Gilliland, R. L. & Hoyt, D. V. *Nature* **300**, 689–693 (1982).
3. Oster, L., Schatten, K. H. & Sofia, S. *Astrophys. J.* **256**, 768–773 (1982).
4. Lawrence, J. K., Chapman, G. A., Herzog, A. D. & Shelton, J. C. *Astrophys. J.* **292**, 297 (1985).
5. Chapman, G. A., Herzog, A. D., Lawrence, J. K. & Shelton, J. C. *Astrophys. J. Lett.* **282**, L99 (1984).
6. Chapman, G. A. *Nature* **308**, 252–254 (1984).
7. Willson, R. C. & Hudson, H. S. *Astrophys. J. Lett.* **254**, L185 (1981).
8. Chapman, G. A. & Groisman, G. *Sol. Phys.* **91**, 45–50 (1984).

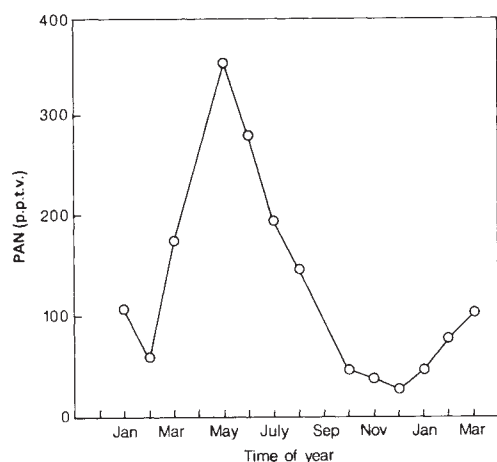


Fig. 1 PAN concentrations between January 1980 and April 1981. $\text{CFC}_{11} < 189$ p.p.t.v. $\text{CH}_3\text{CCl}_3 < 180$ p.p.t.v.

An automatic electron capture gas chromatograph was used to make measurements of PAN, CFC_{11} and CH_3CCl_3 (methyl chloroform). The technique has been described in detail elsewhere^{16,17}; basically a 5-cm³ sample of ambient air is injected into the gas chromatograph every 2 h. Separate injections are made for PAN onto a PEG column and for CFC_{11} and CH_3CCl_3 onto an OV101 column. The sensitivity limit for PAN for this device is <20 p.p.t.v., and of the same order for the two halocarbons. Calibration of PAN was performed as described in ref. 10; CFC_{11} and CH_3CCl_3 were calibrated using a static vapour dilution technique starting from pure compounds¹⁷.

The behaviour pattern of ambient PAN concentration at Harwell has been described in detail elsewhere¹². When all the data are considered, a summer maximum is observed, reflecting the general increase in photochemical activity in polluted atmospheres in the summer months. When the data are selected by the cleanliness of the air, however, a different pattern emerges. PAN in clean air measured at Harwell shows a very pronounced maximum in the spring months, March to May (Fig. 1).

The data shown in Fig. 1 were carefully selected from the full data set by considering only those samples associated with levels of CFC_{11} below a defined maximum. The value chosen for the CFC_{11} concentration (189 p.p.t.v.) was that predicted¹⁶ to be typical of the background atmosphere in 1980–81. Concentrations below this level were required to be observed for at least three consecutive samples between 0900 and 1700 h for the air mass to be regarded as 'clean'. Nighttime values were not used because of the phenomenon observed frequently at Harwell of rapid surface removal from a shallow nocturnal inversion⁸. The data were further screened by selecting only those samples associated with CH_3CCl_3 concentrations <180 p.p.t.v. This molecule is a very sensitive indicator of area contamination and mostly exceeded 200 p.p.t.v. at Harwell during 1980–81. After this careful selection of data, typically 10–20 individual samples per month could be used to study the seasonal variation of PAN in background air (Table 1). There was no obvious correlation between the months with the highest PAN and the highest CH_3CCl_3 . In fact, the spring months with high PAN all showed average CH_3CCl_3 levels below 150 p.p.t.v., whereas the lowest PAN levels were recorded when the CH_3CCl_3 average registered 160 p.p.t.v. We are, therefore, reasonably confident that the high PAN levels recorded in the spring of 1980 are not a purely local phenomenon but probably represent a background average at our location.

If our interpretation is correct then these substantial PAN levels in clean air could indicate that the type of photochemistry responsible for the Los Angeles type of smog is occurring throughout large parts of the Northern Hemisphere troposphere.

Table 1 PAN concentrations at Harwell during 'clean' periods

PAN—Daytime values 0900–1700 (p.p.t.v.)							
	No. of samples	CFCI_3		CH_3CCI_3		PAN	
		$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
1980							
January	11	182	4	172	20	104	30
February	11	184	3	172	26	59	30
March	14	179	6	149	17	178	139
April	—	—	—	—	—	—	—
May	8	181	2	149	13	353	148
June	19	181	4	140	16	283	214
July	8	175	4	—	—	195	130
August	21	182	4	—	—	144	99
September	—	—	—	—	—	—	—
October	6	184	3	—	—	47	36
November	11	182	4	—	—	33	29
December	15	183	4	—	—	27	17
1981							
January	3	185	6	159	20	48	45
February	9	183	2	171	15	78	72
March	38	179	7	182	30	102	73

$\text{CFC}_{11} < 189$ p.p.t.v. and $\text{CH}_3\text{CCl}_3 < 200$ p.p.t.v. for at least 4 h. $\bar{x}$ is the mean and s the standard deviation of the concentration measurements.

This requires the presence of equal quantities of precursor molecules such as hydrocarbons and the oxides of nitrogen. Recent measurements made in the Arctic indicate that the concentrations of many hydrocarbons are large during the months of February and March, and then decline during the summer¹⁸. These hydrocarbons include ethane, acetylene, propane, butanes, pentanes, hexanes, benzene and toluene; all of which are emitted mainly by anthropogenic activity from sources north of 40 °N. In addition, some recent unpublished calculations by J. Fishman suggest a winter maximum for NO_x in the Northern Hemisphere. It is most unlikely that this marked seasonal variation is confined to the Arctic regions and in the case of hydrocarbons, aircraft measurements made by one of us (S.A.P.) in combination with other workers in England, show considerably higher concentrations in clean air in the winter months than in the summer months. The consumption of this hydrocarbon reservoir by photochemistry could certainly produce a spring maximum in the PAN concentration between the winter hydrocarbon maximum and the summer maximum in atmospheric photochemical activity at more northerly latitudes. This should be a large-scale phenomenon which occurs over a wide range of latitude, longitude and altitude. Unfortunately, the only available measurements on seasonal behaviour are the present ones, which are ground-based. However, the few aircraft measurements which have been published show that PAN is often present at substantial concentrations in the free troposphere^{14,15}, suggesting that it is indeed widely dispersed.

The production of PAN on a large scale will slow down as the concentrations of reactants drop during the spring and the PAN concentration will then decay by the processes described earlier. According to Fig. 1, the decay has a half life of the order of 2 months assuming a logarithmic function for the decline in concentration between May and October.

This interpretation of the behaviour of PAN is based entirely on the influence of anthropogenically-emitted hydrocarbons, which are probably emitted at a similar rate throughout the year. Natural emissions of hydrocarbons may also influence the seasonal behaviour, but many of these are very reactive and would not contribute to the winter reservoir. Whatever the source of the precursor molecules, production of PAN is closely associated with production of ozone. Measurements made at Harwell suggest that the simultaneous production ratio is approximately one molecule of PAN to 50 molecules of ozone¹². If the same

production ratio appertains to the 'clean' air conditions then the maximum of 353 p.p.t. of PAN in April and May would be associated with the formation of approximately 15 p.p.b. (parts per 10^9) of ozone. It is quite possible, therefore, that some of the spring increase in ozone concentration observed in the Northern Hemisphere could be due to tropospheric ozone production rather than stratospheric injection. This would be difficult to detect without the coincidental PAN measurements.

More measurements of both these oxidants are urgently required. Also, attempts should be made to contrast the situation in the two hemispheres, since the concentration of non-methane hydrocarbons and NO_x should be lower in the Southern Hemisphere.

This work formed part of a programme of Environmental Research funded by the Department of Environment at Harwell. S.A.P. is currently a visiting scientist at the National Center for Atmospheric Research, Boulder, Colorado 80307. The National Center for Atmospheric Research is sponsored by the NSF.

Received 23 September; accepted 20 December 1985.

1. Penkett, S. A. *Nature* **302**, 293–294 (1983).
2. Demerjian, K. L., Kerr, J. A. & Calvert, J. G. in *Advances in Environmental Sciences*, Vol. 4, (eds Pitts, J. N. & Metcalf, R. L.) 1–262 (Wiley, New York, 1974).
3. Penkett, S. A. in *Atmospheric Chemistry* (ed. Goldberg, E. D.) 329–355 (Springer, Berlin, 1982).
4. Derwent, R. G. & Hov, O. *Harwell Rep. AERE R9434* (HMSO, London, 1979).
5. Levy, H. *Science* **174**, 141–143 (1971).
6. Cox, R. A. & Roffey, M. J. *Envir. Sci. Technol.* **11**, 990–996 (1977).
7. Wallington, T. J., Atkinson, R. & Weiner, A. M. *Geophys. Res. Lett.* **11**, 861–864 (1984).
8. Garland, J. & Penkett, S. A. *Atmos. Envir.* **10**, 1126–1131 (1976).
9. Lee, Y.-N. *Gas Liquid Chemistry of Natural Waters*, Vol. 1, Pap. No. 21 (Brookhaven National Laboratory, 1984).
10. Penkett, S. A., Sandalls, F. J. & Jones, B. M. R. *V.D.I. Berichte*, **270**, 47–56 (1977).
11. Lonneman, W. A., Bufalini, J. J. & Seila, R. L. *Envir. Sci. Technol.* **10**, 374–380 (1976).
12. Brice, K. A. *et al. Atmos. Envir.* **18**, 2691–2702 (1984).
13. Lovelock, J. E. & Penkett, S. A. *Nature* **249**, 434 (1974).
14. Singh, H. B. & Salas, L. H. *Nature* **302**, 326–328 (1982).
15. Meyrahn, H. *et al. Proc. 3rd European Symp. Physico-Chemical Reactions in the Atmosphere* 38–43 (eds Ott, H. & Versino, B.) (Reidel, Amsterdam, 1984).
16. Brice, K. A. *et al. Atmos. Envir.* **16**, 2543–2554 (1982).
17. Penkett, S. A. *et al. Atmos. Envir.* **13**, 1011–1019 (1979).
18. Hov, O. *et al. Geophys. Res. Lett.* **11**, 425–428 (1984).

A regional acidic cloud/fog water event in the eastern United States

Kathleen C. Weathers*, Gene E. Likens*,
F. Herbert Bormann†, John S. Eaton*,
W. Breck Bowden‡, Janet L. Andersen‡, Donald A. Cass‡,
James N. Galloway§, William C. Keene§,
Kenneth D. Kimball||, Paul Huth¶ & Daniel Smiley¶

* Institute of Ecosystem Studies, Millbrook, New York 12545, USA

† Yale School of Forestry and Environmental Studies, New Haven, Connecticut 06511, USA

‡ College of the Atlantic, Bar Harbor, Maine 04609, USA

§ University of Virginia, Charlottesville, Virginia 22903, USA

|| Appalachian Mountain Club, Gorham, New Hampshire 03581, USA

¶ Mohonk Preserve, Inc., New Paltz, New York 12561, USA

Cloud and fog water constitutes an important hydrological input to specific ecosystems^{1–3}. Recently, questions have been raised about the chemical composition of clouds and fog with reference to their potential role in adding chemicals such as nutrients, mineral acids and trace metals to such ecosystems. There are few data on the chemistry of cloud and fog water; those that exist suggest generally low pH values and high concentrations of major inorganic cations, anions and trace metals^{4–7}, especially when compared with rain water collected from the same or nearby locations^{8,9}. Here we present the first analysis of a widespread, episodic cloud/fog event, using samples collected during August 1984 at six non-urban sites in the eastern United States. The pH was extremely low (2.8–3.09) and concentrations of sulphate and nitrate were 7–43 times greater than those for average precipitation at four eastern sites, and higher than previously reported values for cloud/fog water in the eastern United States. This suggests that such water may add ecologically significant amounts of pollutants and nutrients to many ecosystems in the region.

The Cloud Water Project (CWP), initiated in 1983 and coordinated from the Institute of Ecosystem Studies (IES) in Millbrook, New York, was designed as a synoptic study to make inter- and intra-site comparisons of cloud and rain water chemistry across a wide geographical area, ranging from Puerto Rico to Alaska. The 10 CWP sites are located in areas where cloud/fog water (for the purposes of this paper, fog is defined as ground-based cloud) is likely to be an important input of water and chemicals to natural ecosystems, such as coastal and montane regions. During this study, a widespread acidic cloud/fog event was detected in August 1984.

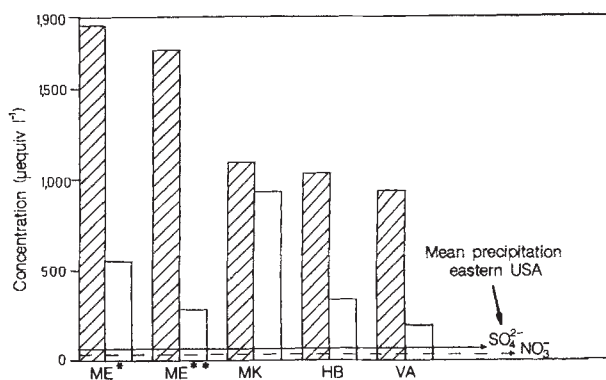


Fig. 1 Sulphate (hatched) and nitrate (white) concentrations of an acidic cloud/fog water event compared with mean precipitation from Whiteface Mountain, New York¹³; Ithaca, New York¹³; Hubbard Brook, New Hampshire¹⁴ and Charlottesville, Virginia¹³. Two-letter site identifiers are as in text. *Sample collected on 7–8 August 1984; ** sample collected on 8–9 August 1984.

During the period 7–13 August 1984, samples of cloud and fog water with field pH values ranging from 2.80 to 3.09 were collected from six non-urban sites (hereafter referred to by site identifiers noted in parentheses): Bar Harbor, Maine (ME); Mt Washington, New Hampshire (LC); Mt Lafayette, New Hampshire (GL); Mohonk Mountain, New York (MK); Hubbard Brook Experimental Forest, West Thornton, New Hampshire (HB) and Loft Mountain, Virginia (VA) (Table 1). Total sampling time at the sites varied from ~3 to 12 h; however, the total duration of the cloud/fog events is unknown.

Cloud/fog collections were made using a CWP Active Collector. This collector uses a battery-powered fan to draw atmospheric droplets through a ventral opening. It was designed so that cloud/fog water collections could be made at both high- and low-wind environs. Droplets are pulled at 7.6 m s^{-1} across Teflon strands, where they are caught (50% efficiency at $5.2\text{-}\mu\text{m}$ drop size) and drip into a polyethylene collection bottle. Cloud/fog water samples were retrieved immediately after collectors were turned off, transported to a field laboratory, measured for pH using a modification of field techniques used by the Global Precipitation Chemistry Project^{10,11} and shipped to IES for chemical analysis.

Sulphate, nitrate, hydrogen and ammonium were the dominant ions in the samples collected from ME, HB, MK and VA, with sulphate and nitrate comprising ~95% of the measured anions, and hydrogen and ammonium ~95% of the measured

Table 1 Location and elevation of cloud water project sites; date, sampling time and pH of acidic cloud/fog event

Site	Elevation (m above mean sea level)	Date	Sampling time (h)	Cloud/fog field pH
Bar Harbor, Maine (ME)	10	7-8 August 1984	1930-0730	2.90
Bar Harbor, Maine (ME)	10	8-9 August 1984	2130-0730	3.00
Bar Harbor, Maine (ME)	10	9 August 1984	0730-1530	2.95†
Mt Washington, New Hampshire (LC)	1,524	10 August 1984	1830-2230	2.97*†
Mt Lafayette, New Hampshire (GL)	1,220	10 August 1984	1830-2230	2.97*†
Mohonk Mountain, New York (MK)	467	11 August 1984	0910-1200	2.80
Hubbard Brook, New Hampshire (HB)	765	12 August 1984	0618-1200	2.97
Loft Mountain, Virginia (VA)	990	13 August 1984	1040-1500	3.09

* Samples from (LC) and (GL) were made by passive collection.

† Sample not of sufficient volume for chemical analyses.

cations. (Cation-anion balances were within $\pm 10\%$.) Organic acids (acetic and formic acids) were measured in the MK sample using techniques described by Keene *et al.*¹⁰. The absolute concentration of organic acids ($50 \mu\text{mol l}^{-1}$) for this cloud water sample was comparable to the highest concentrations reported for individual rain samples collected from central Virginia¹¹. However, the sample was dominated by high concentrations of strong mineral acids, and therefore the percentage of organic acid contribution to total and free acidity was low (3 and $<1\%$, respectively).

Sulphate and nitrate concentrations in cloud and fog water from ME, HB, MK and VA ranged from ~ 930 to $1,850 \mu\text{equiv l}^{-1}$ (sulphate) (The contribution of seasalt sulphate to total sulphate at all sites, calculated from methods described by Keene *et al.*¹² was $<1\%$), and ~ 190 to $900 \mu\text{equiv l}^{-1}$ (nitrate) (Fig. 1). These concentrations are 7-43 times greater than volume-weighted annual mean sulphate and nitrate concentrations in precipitation from locations in the eastern USA. Long-term means from Whiteface Mountain, New York¹³; Ithaca, New York¹³; Hubbard Brook, New Hampshire¹⁴ and Charlottesville, Virginia¹³ range from 44 to $61 \mu\text{equiv l}^{-1}$ sulphate and $21-29 \mu\text{equiv l}^{-1}$ nitrate. With the exception of the nitrate concentration in the VA sample, sulphate and nitrate concentrations of these cloud/fog samples are much higher than cloud water concentrations reported from other sites in New York and New England^{4,6,7} and lie within the range of values reported for urban fogs measured in Bakersfield, California, which are presumably formed in highly polluted conditions⁵.

Before the acidic cloud/fog event, from 3 to 7 August, there was a large, stagnant high-pressure system over the eastern United States. A weakening low-pressure system moved from the north to the south on 7 August, during which time a very slow surface air flow continued to move from the Ohio Valley through the New York-New Jersey corridor and up the eastern seaboard. On approximately 8 August, a cold front stalled off the coast of New England and brought precipitation to much of the eastern United States. By 11 August surface air flow reversed, moving from the north-east to the south-west. The acidic cloud event that followed appears to have covered several thousand kilometres, and enveloped and wetted vegetation at most sites.

Vegetation in the eastern United States is frequently subjected to more than one anthropogenic stress. For example, on 7 August 1984, the day of the first acid fog collection from our Maine site, the State of Maine (Department of Environmental Protection) ozone monitoring station, located ~ 3 km from our ME CWP site, measured six continuous hours of ambient ozone concentrations ≥ 0.09 parts per million (range = $0.090-0.102$ p.p.m.), thereby exceeding the state standard of 0.08 p.p.m. (ref. 15). Apparently, the vegetation surrounding our ME CWP site was subjected simultaneously to an acidic

cloud/fog event and to high levels of ozone. Although the effects of combinations of acidic cloud/fogs and gaseous pollutants are poorly understood, acidic fog/gaseous pollutant combinations are known to produce different effects from those produced by gaseous pollutants alone¹⁶. Researchers have found that many plant species are damaged by ozone concentrations at levels reported for Acadia National Park on 7 August 1984 (ref. 15), and laboratory research has shown that acidic mists of $\text{pH} \leq 3.00$ markedly increase leaching of substances from some vegetation^{17,18}, and that highly acidic mists cause necrotic lesions on the exposed leaves of lettuce¹⁶.

Since our cloud/fog water collections were made by active impact on an artificial surface, we are unable to estimate mass deposition rates of pollutants on surrounding vegetation. However, Lovett *et al.*⁴ have suggested that cloud water deposition in subalpine forests of New England is substantial.⁴ Our data from 7-13 August 1984 suggest that episodic exposure to highly polluted cloud/fog water is characteristic of a broad geographical range of coastal and montane ecosystems in the eastern United States. Additional information from Maine indicates that episodic extreme exposures of mixed pollutants must be carefully considered in evaluating air pollution effects.

To understand more fully the ecological effects of highly acidic cloud/fog episodes such as the event reported above, it is important to document further their frequency and intensity, and their rates of deposition to vegetation.

This project was supported by the Andrew W. Mellon Foundation in a grant to G.E.L. and F.H.B. We thank Bruce Daube, Joni Coury and Wendy Malpass for their support, and Shenandoah National Park, Mohonk Preserve Inc., and the Appalachian Mountain Club for use of facilities.

Received 26 August; accepted 13 December 1985.

- Azvedo, J. & Morgan, D. L. *Ecology* **55**, 1135-1141 (1974).
- Harr, R. D. *Wat. Res. Bull.* **18**, 785-788 (1982).
- Kerfoot, O. *For. Abstr.* **29**, 8-20 (1968).
- Lovett, G. M., Reiners, W. A. & Olson, R. K. *Science* **218**, 1303-1304 (1982).
- Jacob, D. J., Waldman, J. M., Munger, J. W. & Hoffman, M. R. *Tellus* **36B**, 272-285 (1984).
- Castillo, R. A., Kadlecik, J. A. & McLaren, S. *Wat. Air Soil Pollut.* **24**, 323-328 (1985).
- Scherbatskoy, T. & Bliss, M. *Proc. APCA Specialty Meeting: The Meteorology of Acidic Deposition* (Air Pollution Control Association, Pittsburgh, 1984).
- Dollard, G. J., Unsworth, M. H. & Harve, M. J. *Nature* **302**, 241-243 (1983).
- Waldman, J. M. *et al. Science* **218**, 677-679 (1982).
- Lovett, G. M., Galloway, J. N. & Holden, D. J. *Jr J. geophys. Res.* **88**, 5122-5130 (1983).
- Keene, W. C. & Galloway, J. N. *Atmos. Envir.* **18**, 2491-2497 (1984).
- Keene, W. C., Pszenny, A. A. P., Galloway, J. N. & Hawley, M. E. *J. geophys. Res.* (in the press).
- Dana, M. T. & Easter, R. C. *Atmos. Envir.* (submitted).
- Likens, G. E. (ed.) *An Ecosystem Approach to Aquatic Ecology*, 19 (Springer, New York, 1985).
- Bormann, F. H. *Bioscience* **35**, 434-441 (1985).
- Granett, A. L. & Musselman, R. C. *Atmos. Envir.* **18**, 887-890 (1984).
- Scherbatskoy, T. & Klein, R. J. *envir. Qual.* **12**, 189-195 (1983).
- Wood, T. & Bormann, F. H. *Ambio* **4**, 169-171 (1974).

Observation of a dislocation source in ice by synchrotron radiation topography

Shoaib Ahmad, Mutsumi Ohtomo & R. W. Whitworth

Department of Physics, University of Birmingham, Birmingham B15 2TT, UK

A fundamental requirement for the plastic deformation of a crystalline material is a mechanism for the multiplication of dislocations. Such a mechanism has been proposed by Frank and Read¹. Whenever a segment of dislocation lying in its glide plane terminates at a node or at a step taking it out of that plane, the forward motion of the segment on the glide plane causes the dislocation line to spiral around this point, with the result that dislocations can sweep across the plane many times over. There is much evidence for dislocation multiplication in crystals. Sometimes a source in a thin foil throws off a sequence of dislocations while being observed in the electron microscope^{2,3}. In other experiments on bulk specimens the dislocation loops or spirals produced from a source have been observed after the event by the technique of dislocation decoration⁴ or X-ray topography^{5,6}. We present here a sequence of topographs which show a complete source in an ice crystal. The topographs were obtained using the X-ray topography facilities at the Synchrotron Radiation Source Daresbury, UK⁷, and provide an exceptionally clear demonstration of the operation of this kind of Frank-Read source.

The topographs were obtained by allowing a wide but very highly collimated beam of white X rays to fall on the single crystal specimen. Each Laue diffraction spot produced is then an image of the crystal in which dislocations can be seen (provided $\mathbf{g} \cdot \mathbf{b} \neq 0$, where $\mathbf{g}$ is the diffraction vector and $\mathbf{b}$ the Burgers vector)⁸. Photographs were taken of the (1100) image at a Bragg angle of 5.5° corresponding to a wavelength of $0.75 \text{ \AA}$. It was found that high-resolution 35-mm roll film such as Kodalith 2554 gave images at least as good as conventional X-ray films with typical exposure times of $<50 \text{ s}$. The orientation is such that the basal planes on which dislocations glide in ice were projected with little distortion onto the plane of the film. The ice crystal was grown in a glass tube by a Bridgeman technique and had a very low initial dislocation density. The specimen

used was 1.2 mm thick and a region of diameter 10 mm was observed in the topograph. The crystal was maintained at -20°C , and a compressive stress (corresponding to a resolved stress of 0.16 MPa on the basal plane) was applied for periods of 60 s between exposures. Ice is suitable for these observations because it has a relatively low X-ray absorption, allowing the use of thick specimens; it gives dislocation images with good contrast, and the dislocation velocity is not a rapidly varying function of stress⁹.

The topographs reveal the motion of dislocations and a general increase in dislocation density in the very early stages of deformation. One effect of stress is to cause dislocations to lie predominantly in the screw or 60° orientations on their glide plane. Figure 1 shows enlarged portions of six topographs selected from a sequence in which a dislocation source was observed in a region of crystal almost free of other dislocations. The main dislocation segment in *a* is moving from the bottom right to the top left and has developed a cusp marked by the arrow. The arrows in subsequent topographs point to the same absolute point in the crystal. As the source operates, a loop appears to develop from the cusp, a second loop seems to form within the first loop and another begins to form within the second, but these are not actually loops. They are formed by a superposition of two opposite spirals starting from the cusp as shown in the line diagrams to the right of topographs *b*, *d* and *f* in Fig. 1. Our interpretation of these effects is that at the cusp the dislocation undergoes a step from one glide plane to a parallel one a small distance behind. The dislocation on each glide plane then expands as a single-ended Frank-Read source, but the two do not spiral out at quite the same rate.

There is no evidence from the topographs to show how the step was formed. In the position of topograph *a* (Fig. 1), the cusp is on a segment of 60° orientation, so the step cannot have been formed by cross-glide; nor can it be formed by the intersection with another dislocation, as then the glide planes would not be far enough apart for the dislocations on them to cross freely over one another. It may have been formed by climb before the application of the stress, or the dislocation may originate from one grown into the crystal and not lying on a single basal plane.

These observations are part of a larger project on dislocations in ice supported by the SERC, and we are grateful for the use of the facilities at the SERC Daresbury Laboratory.

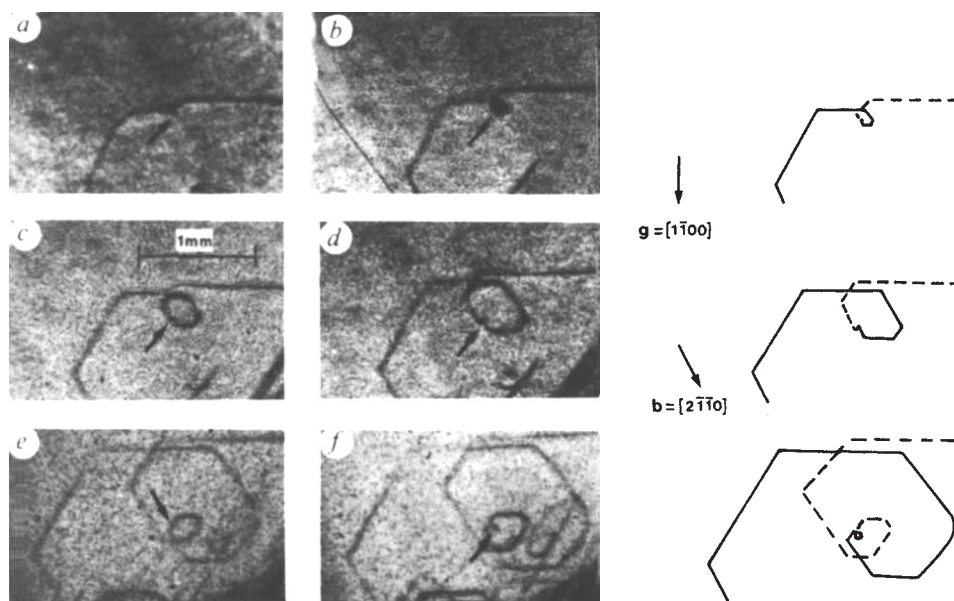


Fig. 1 Six topographs of an ice crystal showing operation of the dislocation source. Line diagrams on the right correspond to topographs *b*, *d* and *f* (top to bottom), with the dislocations shown by solid and broken lines lying on different planes.

Received 9 December 1985; accepted 15 January 1986.

1. Frank, F. C. & Read, W. T. *Phys. Rev.* **79**, 722-723 (1950).
2. Saka, H., Doi, M. & Imura, T. *J. Phys. Soc., Jap.* **29**, 803-804 (1970).
3. Carter, C. B. *Phil. Mag.* **35**, 75-79 (1977).
4. Dash, W. C. *Dislocations and Mechanical Properties of Crystals* (eds Fisher, J. C., Johnston, W. G., Thomson, R. & Vreeland, T.) 57-68 (Wiley, New York, 1957).
5. Authier, A. & Lang, A. R. *J. appl. Phys.* **35**, 1956-1959 (1964).
6. Militat, J. E. A. & Bowen, D. K. *Physica Status Solidi (a)* **3**, 431-445 (1970).
7. Bowen, D. K. *et al. Nucl. Instrum. Meth.* **195**, 277-284 (1982).
8. Tanner, B. K. *X-Ray Diffraction Topography* (Pergamon, Oxford, 1976).
9. Whitworth, R. W. *J. Glaciol.* **21**, 341-359 (1978).

Danish Basin subsidence by Triassic rifting on a lithosphere cooling background

Kai Sørensen*

Geologisk Institut, Aarhus Universitet, DK-8000 Århus C, Denmark

I report here the first quantitative study of the entire post-Palaeozoic subsidence history in the ensemble of north-west European sedimentary basins. Data obtained from wells suggests that the late Permian through Mesozoic subsidence of the Danish Basin is a result of lithosphere-wide cooling following an igneous event of late Carboniferous to early Permian age. This cooling subsidence was superposed by major Triassic rifting and minor mid-Jurassic uplift, both events apparently not having significant thermal consequences. The Danish Basin did not experience the late Jurassic to early Cretaceous rifting and subsequent cooling which were so important in the development of the central North Sea rifts. This result suggests that a reappraisal of North Sea tectonics should look for pre-Jurassic subsidence.

The Danish sedimentary basin north of the Ringkøbing-Fyn High (Fig. 1) contains Mesozoic and Tertiary sediments deposited on Permian evaporites^{1,2}. Late Triassic to Cenozoic salt flow³⁻⁵ induced significant local subsidence anomalies which overprint regional subsidence patterns. To investigate the subsidence history of the basin wells penetrating the Mesozoic sequence have been analysed taking into consideration salt-induced thickness variations. The wells studied (Fig. 1) are located in the deep part of the basin (Mors 1, Skive 1 and Rønde 1) as well as in its shallow parts (Nøvling 1 and Slagelse 1). The Permian evaporites were penetrated in Rønde, Nøvling and Slagelse where they overlie Lower Palaeozoic sediments with an angular unconformity.

In the Rønde and Slagelse wells, no corrections have been made for salt flow in the calculations. The Permian evaporites are relatively thin in both wells (200-300 m) and contain large proportions of anhydrite and dolomite. In Slagelse, they are unstructured and probably represent a marginal facies. Rønde is situated in the deep part of the basin. The anomalously thin evaporite sequence is possibly primary, as the post-Zechstein sediments in Rønde 1 are not significantly thicker than those ~10-20 km to the west where the Zechstein evaporites are of normal thickness (Fig. 1, and I. Madirazza, personal communication). The increasing depth to base Zechstein towards the west of Rønde may, therefore, reflect Zechstein or pre-Zechstein tilting. The Nøvling well is situated close to the northern boundary of the Ringkøbing Fyn High, at the margin of a minor salt pillow. Neither the timing nor the extent of salt layer depletion are well constrained, but seismic profiles⁶ suggest a source layer depletion of between 400 and 600 m. Salt flow occurred during the late Triassic, and continued intermittently into the Cenozoic (see Fig. 1). In the calculations, it has been assumed that 600 m

of source layer depletion occurred during the deposition of the Upper Triassic Tønder and Oddesund Formations. The north-western part of the basin is strongly influenced by salt structuring. For this region, a continuous subsidence curve has been obtained through a combination of data from Mors and Skive. At Mors, salt-layer depletion started during deposition of the Gassum Formation^{5,7}. At Skive, salt flow into a nearby pillow structure had ceased at that time⁸. The subsidence curve for Mors-Skive is a combination of pre-Gassum data from Mors and of younger data from Skive.

The present-day thicknesses indicate Triassic and late Cretaceous subsidence maxima. When subjected to compaction and isostatic corrections, another picture emerges. The compaction correction assumes an exponential decay of porosity (f) with depth (z)

$$f = f_0 \exp(-cz) \quad (1)$$

in which f_0 is the surface porosity. Values of c and f_0 used in this study are shown in Fig. 1. The computational procedure follows ref. 9, using the absolute timescale of ref. 10. As the Permian evaporites overlie lower Palaeozoic sediments with an angular unconformity indicating pre-evaporite uplift, it has been assumed, that they acted as an incompressible basement to the Mesozoic sediments.

The 'driving' tectonic subsidence has been calculated assuming a local, Airy-type isostatic compensation, resulting in the tectonic subsidence curves of Fig. 2, expressed as water-filled basin depths. As the basin fill is characterized by an upwards increasing porosity decay constant (c of equation (1)) the isostatic correction increases with increasing age of the basin. Thus, for example, late Cretaceous tectonic subsidence (157 m) is small compared with the present day thickness (1,228 m) for Mors-Skive. Also, the late Cretaceous tectonic subsidence of Slagelse and Nøvling is not significantly different from that of Mors-Skive, indicating that the geometry of the late Cretaceous basin is not tectonic in nature, but a result of the geometry of the underlying basin.

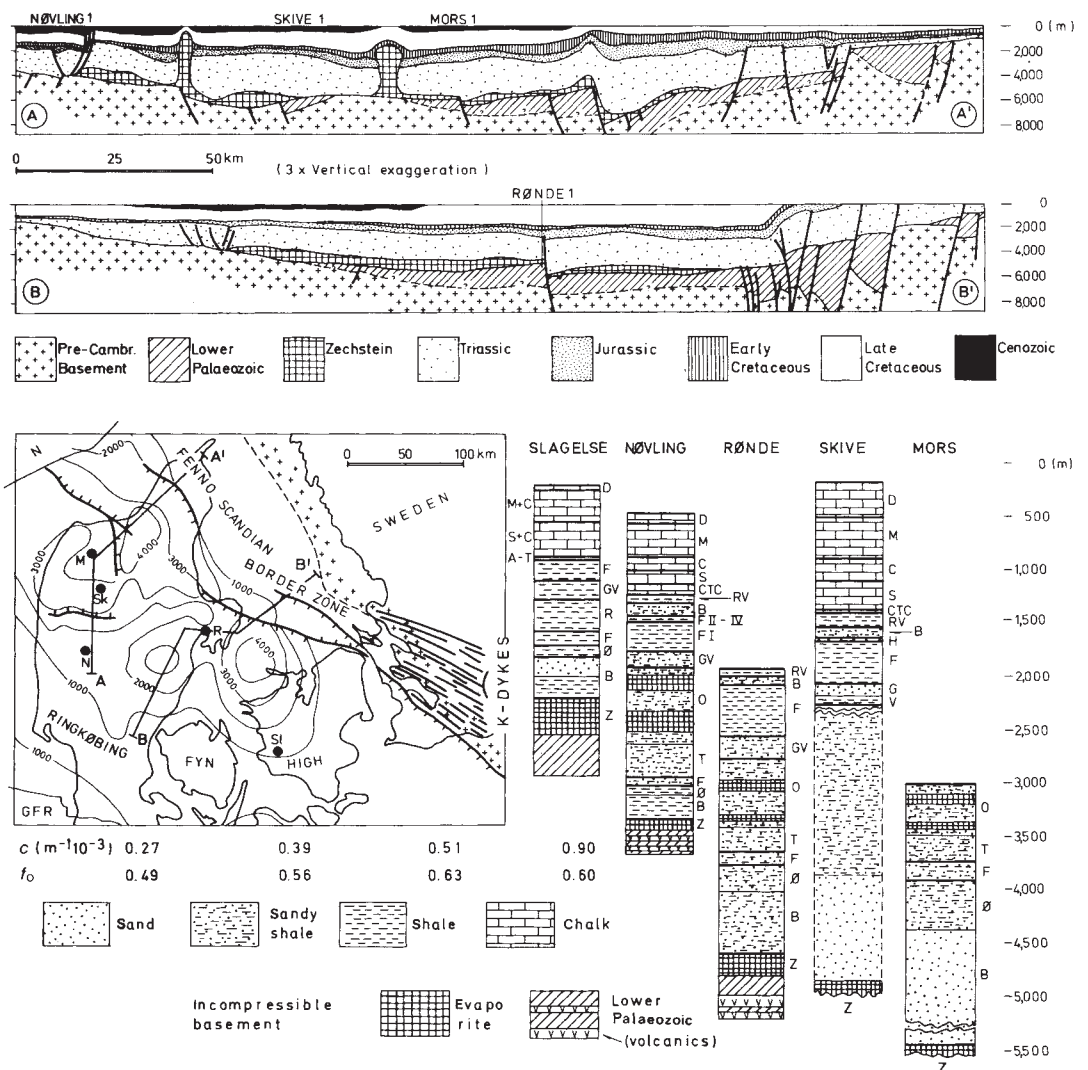
The high subsidence rates during deposition of the Bunter/Ørslev and Oddesund/Tønder Formations, and the magnitude of Triassic subsidence, are akin to rift-controlled subsidence. The subsidence curves for Nøvling, Rønde and Mors-Skive suggest that the Triassic rifting was dual, with the mid-Triassic as an intervening period of relative rest. This impression is in accord with the Triassic subsidence pattern of the neighbouring Polish Basin¹¹.

The notion that sedimentary basins form by extension followed by cooling has recently gained widespread acceptance. McKenzie's theory¹² for instantaneous stretching and one-dimensional heat transfer has been extended to finite rifting times and two-dimensional heat transport¹³. Predictions of this theory are compared with the actual subsidence in Fig. 2 and in Table 1.

For Slagelse, the subsidence curve suggests only one fault-controlled subsidence phase (Bunter/Ørslev). The observed post-Bunter/Ørslev subsidence is significantly larger than the predicted subsidence. This is also the case if 25 Myr extension is assumed. At Nøvling, the observed subsidence is also significantly larger than the predicted subsidence. For Rønde, the observed subsidence magnitude and the predicted magnitude agree for the 25 Myr extension, while for Mors-Skive agreement between observation and prediction is found for the 5-Myr case. Comparison between subsidence curves (Fig. 2) show that the predicted subsidence for both Rønde and Mors-Skive (TrERC of Fig. 2) are systematically faster for the initial 45 Myr cooling than for the observed subsidence. The predictions of Fig. 2 are for marginal positions within a basin (position B of ref. 13), while the geology implies central positions (positions C of ref. 13). In these predictions, the discrepancies between observed and predicted initial 45-Myr cooling subsidence become magnified. We conclude that the subsidence following deposition

* Present address: Statoil a.s., PO Box 300 Forus, N-4001 Stavanger, Norway.

Fig. 1 Danish Basin geometry and analysed wells. The Nøvling 1, Skive 1 and Mors 1 wells lie slightly off-section. Their projected positions appear on profile A-A'. Contours (m) show present-day Triassic thicknesses²³. The Konga dykes¹⁹ of Scania are shown schematically²⁷. Material parameters for decompaction calculations, except for chalk, follow ref. 9. For chalk the decay constant is derived from porosity data on Danish chalk²⁸, while surface porosity is from ref. 29. Stratigraphy and lithology are from refs 1, 6, 23, 24, 30, 31 and lithology and depths of Skive below the depth of termination according to ref. 8. Lithostratigraphy of the pre-Upper Cretaceous in ascending order: Z, Permian evaporites; B, Bunter Formation (in Slagelse Bunter shale and sandstone Formations); Ø, Ørslev Formation; F, Falster Formation; T, Tønder Formation; O, Oddesund Formation; V, Vinding Formation; G, Gassum Formation; F, Fjerritslev Formation; H, Haldager Formation; B, Bream Formation; RV, Rødby and Vedsted Formations. Chronostratigraphy of chalk in ascending order: C, Cenomanian; T, Turonian; C, Coniacian; S, Santonian; C, Campanian; M, Maastrichtian; D, Danian. A-T in Slagelse, Albian to Turonian; R, in Slagelse, 'Ringe beds', a supposed time equivalent to the Tønder and Oddesund Formations of other wells²³. Cross-sections provided by P. A. Ziegler (Maersk Olie og Gas A/S and Shell Olie og Gas Udvinde).



of the Tønder/Oddesund Formations cannot be accounted for by a model involving Triassic extension of a magnitude as required by McKenzie's model.

Instead, I suggest that the subsidence curves are composed of an exponentially decaying subsidence superposed by fast subsidence episodes (in the early and in the late Triassic) and segments representing restricted uplift (mid-Jurassic). These two contributions are not *ad hoc*: Triassic faulting is widespread in the Danish Basin. Most of the faults appearing on Fig. 1 are of Triassic age, although the evaporites located centrally in the basin often cause the faults below the base of Zechstein to be expressed as flexures in the Triassic. The mid-Jurassic uplift segments are contemporaneous with uplift in most of the North Sea^{14,15} (mid-Cimmerian uplift). If these segments are omitted, the remaining parts of the subsidence curves can be linked together (Fig. 3) to fit an exponential subsidence with time (t)

$$h = h_0 \exp(-t/a) \quad (2)$$

where h is the height above final subsidence level, a is the time constant and h_0 is the total subsidence.

No corrections have been made for depth of deposition and eustatic sea level to the subsidence curves of Fig. 2, as required for complete backstripping¹⁶. Such corrections may be significant for the assessment of total subsidence, h_0 . Extreme values of h_0 result from two different interpretations of the Cretaceous subsidence. Assuming that depth of deposition and

eustatic corrections are insignificant to the Cretaceous subsidence curve, then the exponentially decaying subsidence must have already reached an asymptotic level during the early Cretaceous. In this case, minimum values of h_0 and the time constant, a , result (Fig. 3), and the late Cretaceous chalk sedimentation requires a tectonic impulse (during the Santonian of Mors-Skive, Fig. 2).

On the other hand, assuming that the depth of deposition increased before the onset of chalk sedimentation, subsidence at an asymptotic rate occurs during the late Cretaceous, yielding an upper limit to the total subsidence, h_0 , and hence to the time constant, a , (Fig. 3). Increased depth of deposition may have resulted from burial of source regions during an early Cretaceous sea-level rise^{10,17}. In this case, the chalk sedimentation may have started by filling a pre-existing basin only ~200 m deep, then continued by isostatic compensation fuelled by rapid (high c) chalk compaction. The change to chalk sedimentation can then be envisaged as non-tectonic, resulting, for example, from the onset of open sea circulation.

The data can be easily fitted to both models. The significant factor, at the present, admittedly preliminary, level of understanding is the magnitude of the time constant for the deep wells Rønde and Mors-Skive, from which lithosphere wide cooling seems an inescapable inference. The ~50-Myr time constant magnitude derived here compares well with time constants derived from cooling-induced, post-accretion subsidence of the

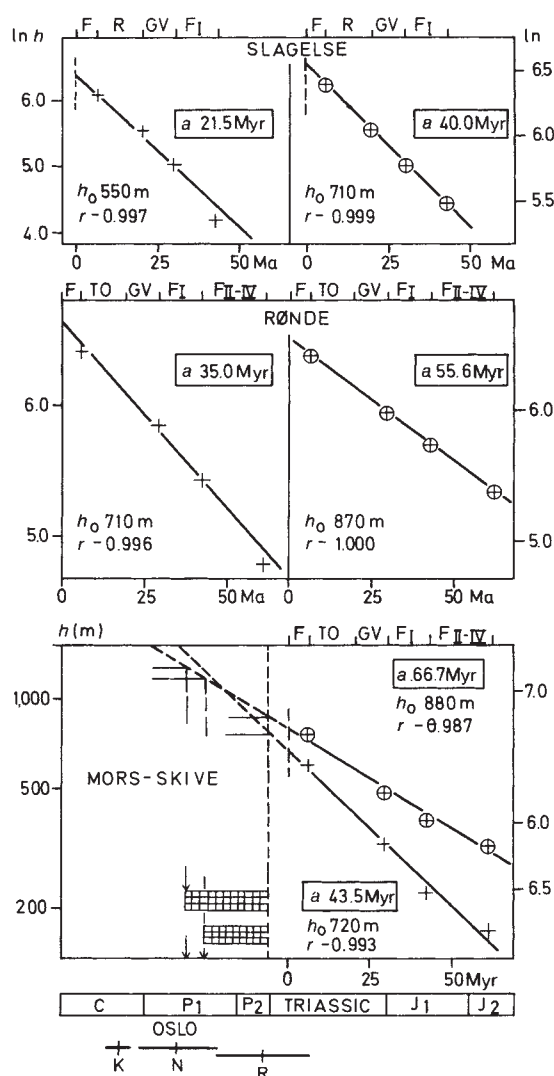


Fig. 2 Tectonic subsidence curves. Stratigraphical symbols along upper margin as in Fig. 1. The stratigraphical extent of the 'gap' created by the mid-Jurassic uplift increases upwards in the diagram. TrERC: the subsidence predictions for cooling following the Triassic (Bunter, B, to Oddeund, O, formation times) extension and rifting according to ref. 13 for marginal basin positions (position B of ref. 13, see Table 1), giving the best agreement between predictions and observations (see text).

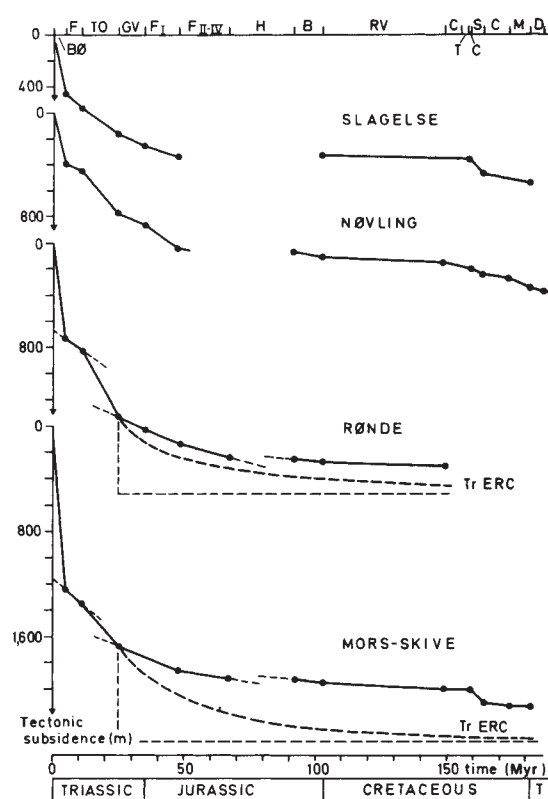


Fig. 3 Exponentially decaying subsidence at Slagelse, Rønde and Mors-Skive. The curves combine Falster, Vinding, Gassum and Fjerritslev formation data, taking $t = 0$ equal to the start of Falster Formation deposition (mid-Triassic). The total subsidence (h_0) was derived as follows: as a first approximation the Tønder-Oddesund and mid-Jurassic 'gaps' were linked by linear extrapolations (stippled lines in Fig. 2). The best exponential fit to this first approximation was then derived. Then the mid-Jurassic subsidence was calculated by an exponential extrapolation across the mid-Jurassic gaps. The lower bounds to h_0 assume the subsidence to have decayed to asymptotic values during the Lower Cretaceous, while the upper bound to h_0 assumes the asymptotic subsidence to be approached during the Upper Cretaceous. For Mors-Skive, the maximum stratigraphical age of the Permian evaporite basin, assuming it to be part of the cooling subsidence, is predicted for both cases (cross-hatched bars). Isotopically-determined ages for Konga dykes²⁰ (K), Nøvling²¹ (N) and Rønde²¹ (R) volcanics and Oslo igneous province²² are shown below the Mors-Skive diagram.

Table 1 Actual subsidence and extension predictions (in m) according to Fig. 6 of ref. 13

	5 Myr extension				25 Myr extension			
	Rifting observed	predicted	Cooling observed		Rifting observed	predicted	Cooling observed	
Slagelse	450	339 (B)	739-899		760	326 (B)	429-589	
Nøvling	400	301 (B)	755-962		760	326 (B)	374-582	
Rønde	720	720 (C)	1,210-1,380		1,327	596 (B) 813 (C)	621-781	
Mors-Skive	1,230	1,230 (C)	1,050-1,210		1,677	753 (B) 1,027 (C)	608-768	

The italic numbers are the predictions for the positions considered in best accordance with the geology. B, Marginal basin position; C, central basin position. The lower value for the range of observed subsidence corresponds to asymptotic decline during the early Cretaceous, while the upper value corresponds to asymptotic decline during the late Cretaceous.

oceanic lithosphere¹⁸. The lower time constant at Slagelse indicates, that the margin of the basin is a result of a southeastwards shallowing thermal anomaly. It remains to be decided where to anchor this cooling subsidence.

It is tempting to interpret the Permian evaporites as a response to this cooling subsidence, as this interpretation provides the necessary subsidence and a reason for the complete lack of terrigenous influx to the evaporite basin. Such an interpretation of the North European Zechstein basins has already been suggested^{14,15} but does not seem to be supported by the data, as the time necessary for a subsidence equal to the evaporite thickness in, for example, Mors exceeds the Zechstein duration by 10–15 Myr (Fig. 3). On the other hand, the Zechstein thickness at Rønne (~100 m, isostatically corrected) requires less time than the duration of the Zechstein. These findings can be reconciled with a cooling subsidence interpretation if it is assumed that evaporite deposition started when the basin was already established and up to ~200 m deep. This interpretation implies an anchoring of the cooling curve somewhere within the earliest Permian or before. For this anchoring there can be provided supporting evidence. In Scania, where a part of the basin has been uplifted, a dolerite dyke swarm, the Konga dykes¹⁹ (Fig. 1), of age 295 ± 10 Myr (ref. 20), trends parallel to the basin, with the northeastern boundary roughly coinciding with the northeastern deep-shallow transition of the Danish Basin. Furthermore, volcanic rocks in Nøvling and Rønne have yielded Permian K–Ar ages²¹. They have been interpreted as Silurian extrusives⁶ due to the age of their enclosing sediments, isotopically reset by a Permian heating event²¹. Alternatively, they may be interpreted as Permian sills. Whatever their true age of intrusion is, they suggest significant Permian heating, directly or indirectly. Finally, the Oslo igneous province, of early Permian age²², as well as the widespread occurrence of Stephanian–Autunian volcanics elsewhere in north-west Europe^{14,15}, points to a significant, regional heating event.

There is, therefore, evidence for regional heating at the time preceding the first evaporite sedimentation. The evidence for heating and cooling, therefore, support my conclusion that the Danish Basin formed by lithosphere-wide cooling beginning during the early Permian. There is no direct evidence for Permo-Carboniferous tectonism in the Danish Basin, except for the extension associated with the intrusion of the Konga dykes (6–7% where exposed in Sweden; R. Stanfors, personal communication). It may be surmised, however, that the pronounced tilting of the pre-Zechstein which is observed in Danish wells is a result of this tectonism.

I envisage the post-earliest Permian history of the Danish Basin as: cooling eventually leading to subsidence below sea level of a plateau which was probably capped by volcanics, initially, then removed during early Permian heating induced uplift. Subsidence below sea level may have preceded evaporite deposition by a few million years. At this time, the brittle crustal thickness must have been at a minimum with optimum conditions for the activation of crustal faults. Onset of rifting at the Permo-Triassic transition may reflect this weak crust, leading to the isostatic sinking of a fault-bounded basin, balanced by uplift of a nearby, probably northeastern^{2,23,24} source area, supplying the basin with terrigenous sediments. Further cooling led to stabilization of the crust preventing further major faulting. This cooling subsidence was interrupted by mid-Jurassic uplift which does not seem to have influenced the cooling trend significantly.

I conclude that the occurrence of Permian evaporites in the entire North European basin indicates a shared, early subsidence history of which the Danish Basin may be an example. I suggest that the subsidence studies made so far of the late Mesozoic and Cenozoic history of the central North Sea basin^{9,25,26} should be augmented by studies of pre-Jurassic subsidence.

I thank L. Jans and B. Ø. Hansen for technical assistance with the manuscript and figures, and Olaf Micheelsen for critical reading of an early manuscript, also I. Madirazza for helpful

discussions of several geological problems and P. A. Ziegler for a helpful review and efforts leading to the release of, and incorporation into the manuscript of the profiles of Fig. 1.

Received 9 July; accepted 19 November 1985.

1. Sorgenfrei, T. & Buch, A. *Geol. Surv. Denmk III Rk.*, 36 (1964).
2. Larsen, G. *Geol. Surv. Denmk II Rk.* 91 (1966).
3. Madirazza, I. *Bull. geol. Soc. Denmk* 24, 161–171 (1975).
4. Madirazza, I. & Pedersen, S. A. Preprint (in preparation).
5. Sørensen, K. *Nature* 319, 23–27 (1986).
6. Rasmussen, L. B. *et al. Geol. Surv. Denmk III Ser. No. 40* (1973).
7. Larsen, G. & Baumann, J. *Dansk geol. Foren., Årsskr. f. 1981*, 151–155 (1982).
8. Møller, J. J. thesis, Aarhus Univ. (1982).
9. Sclater, J. G. & Christie, P. A. F. *J. geophys. Res.* 85, 3711–3739 (1980).
10. Harland, W. B. *et al. A Geologic Time Scale* (Cambridge University Press, 1982).
11. Sokolowski, S. (ed.) *Geology of Poland* 1, Pt. 2, Warsaw (1976).
12. McKenzie, D. P. *Earth planet. Sci. Lett.* 40, 25–32 (1978).
13. Cochran, J. R. *Earth planet. Sci. Lett.* 66, 289–302 (1983).
14. Ziegler, P. A. in *The Petroleum Geology of the Continental Shelf of North-west Europe* (eds Illing, L. V. and Hobson, G. D.) 3–39 (Heyden, London, 1981).
15. Ziegler, P. A. *Geological Atlas of Western and Central Europe* (Elsevier, Amsterdam, 1982).
16. Steckler, M. S. & Watts, A. B. *Earth planet. Sci. Lett.* 41, 1–13 (1978).
17. Vail, P. R., Mitchum, R. M. & Thompson, S. *Am. Ass. petrol. Geol. Mem.* 26, 83–97 (1977).
18. Parsons, B. & Sclater, J. G. *J. geophys. Res.* 82, 803–827 (1977).
19. Törnebohm, A. E. K. *Sven. Vetenskapsakad. Hand.* 14 13 (1877).
20. Klingspor, I. *Geol. Fören. Stockh. Förh.* 98, 195–216 (1976).
21. Larsen, O. *Dansk geol. Foren., Årsskr. f. 1971*, 91–94 (1972).
22. Sundvoll, B. in *Petrology and Geochemistry of Continental Rifts* (eds Neumann, E.-R. & Ramberg, I.) (Reidel, Dordrecht, 1978).
23. Berthelsen, F. *Geol. Surv. Denmk, Ser. B No. 4* (1980).
24. Michelsen, O. *Geol. Surv. Denmk, Ser. B No. 2* (1978).
25. Wood, R. J. *Earth planet. Sci. Lett.* 54, 306–312 (1981).
26. Barton, P. & Wood, R. *Geophys. J. R. astr. Soc.* 79, 987–1022 (1984).
27. Bergström, J. *et al. Geol. Surv. Sweden, Ser. Ca No. 54* (1982).
28. Geological Survey of Denmark, *Hydrogeological Programme, Mors Salt Dome: Geological Main Rep.* (1981).
29. Schmoker, J. W. *Am. Ass. petrol. Geol. Bull.* 68, 1697–1703 (1984).
30. Rasmussen, L. B. *et al. Geol. Surv. Denmk III Ser. No. 39* (1971).
31. Michelsen, O. *et al. Geol. Surv. Denmk, Ser. B No. 5* (1981).

Implication of welded breccia in Muong Nong-type tektites

Darryl S. Futrell

6222 Haviland, Whittier, California 90601, USA

Tektites are considered to have originated (1) as ejecta from terrestrial cometary impacts^{1,2} or from terrestrial meteorite or asteroid impacts^{3,4}; (2) in cometary impacts in which splash-form tektites were ejected and Muong Nong-type tektites were produced in their place^{5,6}; or (3) as volcanic ejecta from the Moon⁷. Welded breccias (Fig. 1) have recently been reported in several Muong Nong-type tektites from Thailand⁸. Two additional welded breccia structures of a more complex nature are described here which show evidence of having experienced two distinct periods of high temperatures during their formation, separated by a minimum of several minutes of cooling through the annealing range. This implies a prolonged multi-stage formation process entirely at the site of origin. In contrast, theories of ejection of tektite glass from terrestrial meteorite impact craters or cometary collision events imply instant formation.

Muong Nong-type tektites are found as layered slab-like chunks and fragments ranging in size up to a specimen 19 cm thick normal to the layering, weighing 12.8 kg (ref. 6). These tektites have the same major-element composition as the splash-form tektites from the same Australasian strewn-field⁹, but differ in exhibiting a strong enrichment in volatiles other than water^{10–14}.

Following a report of angular and pointed bubbles in Muong Nong-type tektites¹⁵, O'Keefe and Adler¹⁶ showed that a Muong Nong-type tektite was composed of separate sub-millimetre rounded particles of glass that were pressed and welded together while soft (Fig. 2). Walter and Cassidy¹⁷ described Muong

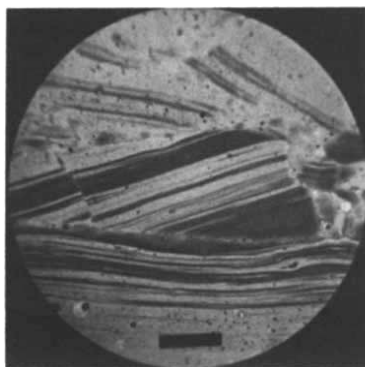


Fig. 1 Photomicrograph in transmitted light of a 250- μ m thick section of breccia welded to a typically layered 87.0-g Thailand Muong Nong-type tektite. Scale bar, 1 mm.

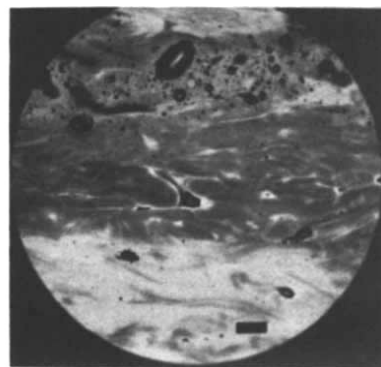


Fig. 2 Photomicrograph in transmitted light of a 300- μ m thick section of a Muong Nong-type tektite from Kan Luang Dong, Thailand. In the centre, angular voids (black) partially outline dark-coloured lenticules welded together while soft. Scale bar, 200 μ m.

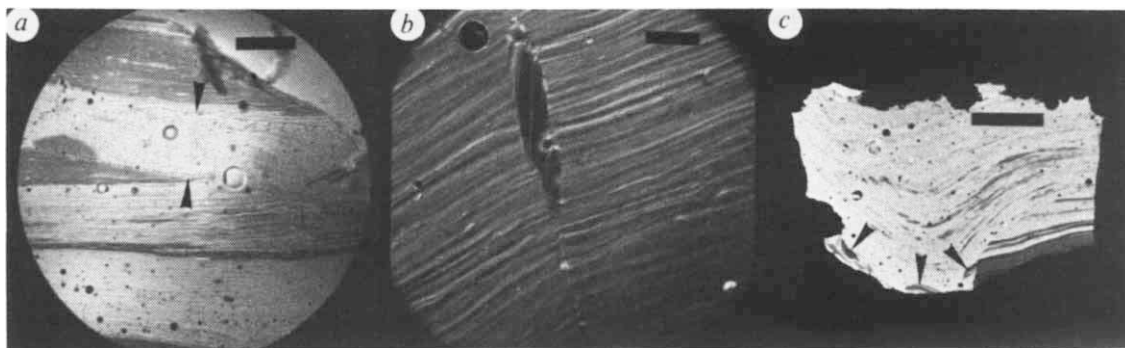


Fig. 3 *a*, Photomicrograph in occulted transmitted light of a 250- μ m thick section containing a bed of buried welded breccia (thickness indicated by arrows) in a Thailand Muong Nong-type tektite of 44.8 g. Dark marks in upper right quadrant along edge of specimen are artefacts on the glass slide. USNM no. 6246-1. Scale bar, 1 mm. *b*, Photomicrograph in occulted transmitted light of a breccia fragment trapped in a vertical welded fault 8 mm below the buried bed of welded breccia in *a*. USNM no. 6246-1. Scale bar, 200 μ m. *c*, Photomicrograph in transmitted light of a 200- μ m thick section of a portion of a V-shaped fissure excavated by brecciation and filled with an additional deposit of glass in a Thailand Muong Nong-type tektite of 30.6 g. Arrows indicate breccia fragments in the trough of the fissure. The left wall of the fissure was lost in sectioning. Scale bar, 5 mm.

Nong-type tektites as being "...composed of relatively homogeneous grains of glass, lenticular in shape and from 50 to 300 μ m in short dimension...". In most specimens, however, these glass grains are not recognizable because they have been drawn out into thin glassy threads due to flow¹⁵. The term 'lenticules', defined as small lens-shaped bodies in a rock mass¹⁸, has been suggested for these rounded glass particles⁷ and will be used here. It has been demonstrated that the lenticules do not have the compositions of any distinct minerals, instead, they all have the composition of homogeneous glasses, at least qualitatively like the bulk composition of the tektite¹⁹.

The above studies demonstrate that Muong Nong-type tektites are welded accretions of lenticules. These accretions occasionally also contain sparsely scattered, sub-millimetre, gravitationally sorted relict mineral grains²⁰; coesite grains inside sub-millimetre silica inclusions have also been reported²¹.

Two very significant configurations of welded breccia in Muong Nong-type tektites from the Ubon Ratchathani region of Thailand have now been observed. One, a specimen of 44.8 g, contains a buried bed of welded breccia with an average thickness of 1.2 mm (Fig. 3*a*). The contact on one side of this bed cuts diagonally through several dark- and light-coloured layers. This layering has a total thickness of at least 1.5 cm. The contact on the opposite side of the breccia consists of undisturbed

layered glass at least 3.5 mm thick. Note also a breccia fragment trapped in a welded fault (Fig. 3*b*).

The other specimen, of 30.6 g, contains a 'V'-shaped fissure at least 1 cm deep apparently spalled out of near-planar layered glass, with a thickness normal to the layering of at least 1.5 cm (Fig. 3*c*). This fissure contains additional glass with banding that follows its contours and those of the glass remaining on either side of the fissure to a depth of at least 5 mm. A few scattered breccia fragments are visible in the trough of the fissure along the contact between its walls and the layering following its contours.

An estimate is needed of the time required for slabs of tektite glass of similar thickness to cool from their accretion temperatures to a temperature at which the fastest-cooling portion of the slabs would be quite brittle and capable of brecciating to the degree observed.

A temperature near 2,000 °C has been suggested for the formation of tektites in an impact^{22,23}, and 1,200 °C as the volcanic gas and magma temperature in the lunar volcanic origin hypothesis⁷. Although the lenticules that make up the layered tektites must have experienced higher temperatures during their original genesis, the angular and pointed bubbles found in many Muong Nong-type tektites, representing pore space between lenticules that barely welded, indicate accretion and welding

temperatures slightly less than 900 °C, since this is the temperature at which they become perfect spheres due to surface tension²⁴. The lack of angular and pointed bubbles in these breccia-containing specimens indicates accretion temperatures in excess of 900 °C. Therefore, an estimated formation temperature of 1,000 °C will be used here. On the other hand, 500 °C appears to be the maximum temperature permitting brecciation of cooling tektite glass, since it is below both the annealing point of 650 °C (ref. 25) and the strain point of 600 °C (ref. 26).

As there are no studies of the cooling rate of tektite glass slabs, estimates must be made based on the work with a cut Muong Nong-type tektite glass sphere of 2.5-cm radius by Klein *et al.*²⁷. Let us suppose the cooling slab to be radiating heat from the top surface, and to be losing very little, if any, heat at the bottom, because it must rest on some, probably insulating, surface. Since the ratio of exposed surface to volume is greater for a sphere of radius a than for a slab of thickness a with just one side exposed, the slab will take longer to cool at its surface and at depth. Therefore, the elapsed cooling times for the sphere can be considered as minima for the slab.

It is found by interpolation that ~8 min would be required for the surface of the sphere of 2.5 cm radius to cool from 1,000 to 500 °C; for the centre 11.5 min would be required. Most of the specimens with welded breccia do have thicknesses near 2.5 cm (ref. 8). Although the two specimens described here are not that thick, they have certainly lost material through natural solution-etching, and possibly through abrasion and/or ablation. Thus, an extended period of several minutes of cooling, rather than seconds, is required before brecciation can occur.

There are several clues to help determine which side is uppermost. In every case, except for a specimen containing breccia trapped within an isoclinal fold⁸, on one side of the breccia bed the contact is always irregular, cutting diagonally through several layers, and in some cases the unbreciated glass becomes more broken, beginning with microfaults, as the bed of breccia is approached. The breccia fragment trapped in a welded fault provides additional evidence that this side of the breccia bed in all cases is the underside and hence precedes the breccia. In the case of the buried breccias, on the opposite side of the breccia bed (that is, the top of the specimens and undisturbed except for slight flow), layers of glass have filled in around the breccia and hence must occur after it.

The layered glass both underlying and overlying the welded breccia in the two specimens described demonstrates that both were subject to the same lenticule accretion process both before and after the cooling and brecciating episode of several minutes. The following sequence of events must have occurred during their formation in order to account for the structures described: (1) The formation of quantities of sub-millimetre-lenticules of tektite glass, followed by their deposition or accretion in layers on some surface. (2) The welding of these layers. (3) A pause of several minutes in the accretion process allowing sufficient cooling to permit partial brecciation involving the then exposed top surfaces, probably as a result of stress from flow in the then still partially soft underlying glass. (4) A resumption of high temperatures of at least 900 °C, accompanied in these two cases by additional deposits of lenticules, enveloping the loose breccia fragments and fusing them to the underlying intact portions. (5) An additional period of cooling. (6) Eventual fragmentation into individual specimens and natural distribution to present sites.

The formation of this type of tektite requires an extended period of time, in some cases measured in tens of minutes. To consider the time available in an impact, I note that 10 km has been suggested as a minimum diameter of an impact crater theoretically capable of producing tektites²⁸⁻³⁰. The 18-km-diameter Elgygytgyn structure in eastern Siberia is the largest extant crater proposed as a source for the Australasian tektites³¹. Theoretical considerations by Gault *et al.*³² suggest a first-order estimate of ~25 s for the basic structure and morphological

features to be formed in an impact crater of 18-km-diameter. Even if there were an ejection mechanism somehow capable of ejecting tektite glass as late as 25 s after impact, Muong Nong-type tektites at that point would at best just be beginning to cool, several minutes before being capable of brecciation.

All Muong Nong-type tektites, as accretions, must have remained at their original point of formation, resting on some surface, throughout their entire formational sequence. However, for them to have been produced in place on Earth during the impact of a comet would require an impact site at least 2,400 km in diameter. This is the distance between Muong Nong-type tektite sites that are part of the same Australasian strewn-field event^{33,34} in Thailand⁹ and in the Philippines³⁵, and is about 100 times larger than the diameters of comet nuclei⁷. The coma of a comet is much larger, but Delsemme³⁶ concluded that it is so tenuous that no mechanical effects of a collision with it are likely to be felt at the Earth's surface.

The welded breccias described here and previously⁸ are all very similar to volcanic flow breccias, though on a minute scale. Volcanic flow breccias form from the fragmentation of lava by further movement after part of the mass has congealed³⁷. Welded flow breccias with all the characteristics described here would not be expected from an impact, and as far as I am aware, none have been reported.

There is a remarkable megascopic flow structure similarity between this type of tektite and terrestrial volcanic glass³⁸. If, however, Muong Nong-type tektites represent a volcanic process, it is not terrestrial. All tektites differ from terrestrial igneous rocks in that they are much dryer³⁹ and have consistent chemical differences⁴⁰.

I conclude that the welded breccia found in Muong Nong-type tektites provides physical evidence demonstrating a time-consuming multi-stage formation occurring entirely at the point of lenticule accretion. The minimum of several minutes required for these complex structures to be formed precludes an origin involving ejection from a meteoritic impact crater or a cometary collision event.

Received 30 July; accepted 19 December 1985.

1. Urey, H. C. *Nature* **179**, 556-557 (1957).
2. Lin, S. C. *J. geophys. Res.* **71**, 2427-2437 (1966).
3. Spencer, L. J. *Nature* **131**, 117-118 (1933).
4. Cohen, A. J. *COSPAR 3rd int. Space Sci. Symp.* 950-973 (1963).
5. Urey, H. C. *Nature* **197**, 228-230 (1963).
6. Barnes, V. E. *Chem. Erde* **30**, 13-19 (1971).
7. O'Keefe, J. A. *Tektites and their Origin* (Elsevier, Amsterdam, 1976).
8. Futrell, D. S. & Fredriksson, K. *Meteoritics* **18**, 15-18 (1983).
9. Barnes, V. E. & Pitakpaivan, K. *Proc. natn. Acad. Sci. U.S.A.* **48**, 947-955 (1962).
10. Müller, O. & Gentner, W. *Meteoritics* **8**, 414-415 (1973).
11. Moore, C. B. *et al. J. non-cryst. Solids* **67**, 345-348 (1984).
12. Koerber, C. *et al. Lunar planet. Sci.* **15**, 441-442 (1984).
13. Koerber, C. *et al. Lunar planet. Sci.* **15**, 443-444 (1984).
14. Koerber, C. *et al. Lunar Planet. Sci.* **15**, 445-446 (1984).
15. Barnes, V. E. *Geochim. cosmochim. Acta* **28**, 1267-1271 (1964).
16. O'Keefe, J. A. & Adler, I. *Proc. Caltech-JPL lunar Planet. Conf.*, 92-101 (1966).
17. Walter, L. S. & Cassidy, W. *Geol. Soc. Am. 1967 A. Meet.*, New Orleans, 231-232 (1968).
18. Fay, A. H. *U.S. Bur. Mines Bull.* **95**, (1948).
19. Walter, L. S. *NASA TM X62593* (Goddard Space Flight Center, Greenbelt, Maryland, 1968).
20. Glass, B. P. *Earth planet. Sci. Lett.* **16**, 23-26 (1972).
21. Walter, L. S. *Science* **147**, 1029-1032 (1965).
22. Cohen, A. J. *Nature* **188**, 653-654 (1960).
23. Chao, E. C. T. in *Tektites* (ed. O'Keefe, J. A.) 51-94 (University of Chicago Press, 1963).
24. Barnes, V. E. & Russell, R. V. *Geochim. cosmochim. Acta* **30**, 143-152 (1966).
25. Hubbard, D. *et al. EOS* **37**, 767-778 (1956).
26. Hammond, C. R. *Pop. Astr.* **58**, 345-350 (1950).
27. Klein, L. C. *et al. J. geophys. Res.* **85**, 5485-5489 (1980).
28. Hawkins, G. S. *J. geophys. Res.* **68**, 895-910 (1963).
29. Hawkins, G. S. *et al. Geochim. cosmochim. Acta* **28**, 1011-1014 (1964).
30. Taylor, S. R. *Chem. Geol.* **4**, 451-459 (1969).
31. Dietz, R. S. *Meteoritics* **12**, 145-157 (1977).
32. Gault, D. E. *et al. in Shock Metamorphism of Natural Materials* (eds French, B. M. & Short, N. M.) 87-99 (Mono Book Corporation, Baltimore, Maryland, 1968).
33. Chapman, D. R. *et al. Geochim. cosmochim. Acta* **28**, 821-839 (1964).
34. Chapman, D. R. *J. geophys. Res.* **76**, 6309-6338 (1971).
35. Chapman, D. R. & Scheiber, L. C. *J. geophys. Res.* **74**, 6737-6776 (1969).
36. Delsemme, A. H. in *Tektites* (eds Barnes, V. E. & Barnes, M. A.) 358 (Dowden, Hutchinson & Ross, Stroudsburg, 1973).
37. Fisher, R. V. *Bull. geol. Soc. Am.* **71**, 973-982 (1960).
38. Futrell, D. S. *Meteoritics* **12**, 232-235 (1977).
39. Gilchrist, J. *et al. J. geophys. Res.* **74**, 1475-1483 (1969).
40. Suess, F. E. *Mitt. geol. Ges. Wien* **7**, 51-121 (1914).

A new synthetic approach to the ferritin core uncovers the soluble iron(III) oxo-hydroxo aggregate $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$

Sergiu M. Gorun & Stephen J. Lippard

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Hydrolytic polymerization of iron(III) occurs in many reactions *in vivo*, for example, the formation of bacterial magnetite in magnetotactic organisms¹, biomineralization of iron² and the synthesis of the metallic core of the iron-storage protein ferritin³. The ferritin core contains aggregates of up to 4,500 oxygen-bridged, octahedrally coordinated, high-spin iron(III) centres and is attached to the protein shell through carboxylate groups of amino-acid side chains. The X-ray and electron-diffraction patterns of this core resemble those of the mineral ferrihydrite⁴, a hydrated iron oxide formed in nature, *inter alia*, by iron-dependent bacteria. The preparation and structural characterization of such large poly-iron aggregates has been a challenge to inorganic chemists^{5,6}. We have recently shown that tri- and tetranuclear iron(III) oxo complexes of the type thought^{3,7,8} to be important in ferritin-core formation can be prepared by reacting mononuclear $[\text{FeCl}_4]^-$ and binuclear $[\text{Fe}_2\text{OCl}_6]^{2-}$ components in aprotic solvents (ref. 9 and S.M.G., W. H. Armstrong and S.J.L., in preparation). Here we report the discovery of a remarkable new molecule, $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$, obtained by hydrolysis of the $\{\text{Fe}_2\text{O}\}^{4+}$ unit in the presence of limited amounts of water and carboxylate salts. The synthesis and properties of this soluble iron(III) oxo-hydroxo aggregate should help to elucidate the mechanism of formation of poly-iron centres.

$[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$ was prepared either by vapour diffusion of water or by liquid diffusion of aqueous tetrahydrofuran (THF) into acetonitrile solutions of $(\text{Et}_4\text{N})_2[\text{Fe}_2\text{OCl}_6]$ (ref. 10) and $\text{Na}(\text{O}_2\text{CPh})$ to form red-brown triclinic or rhombohedral crystals, respectively (Table 1). More rapid addition

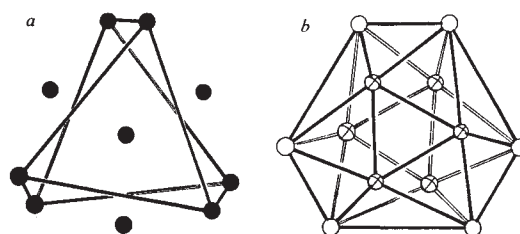


Fig. 1 Structural properties of $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$. *a*, Arrangement of iron atoms viewed down the pseudo- C_3 axis in the triclinic form, which is a crystallographically required C_3 axis in the rhombohedral form. The twist angles are 16.2, 16.9 and 19.9° for the triclinic and 17.4° for the rhombohedral forms. They are defined as the angles between the three pairs of iron atoms forming the triangular faces and the geometric centre of the trigonal prism, as projected onto the plane perpendicular to the (pseudo) 3-fold symmetry axis. *b*, Arrangement of oxygen atoms. Spherical crosses, μ_3 -oxo groups; open circles, μ_3 -hydroxo groups. Selected geometric information for the triclinic and rhombohedral forms follow, respectively: average Fe—O(oxo) bond length, 1.91 and 1.91 Å; Fe—O(hydroxo), 2.11 and 2.09 Å; Fe—O(benzoato), 2.07 and 2.08 Å. Sum of Fe—O—Fe angles around μ_3 -oxo, 360 and 360°; around μ_3 -hydroxo, 312° and 312°.

or greater amounts of water led to the formation of flocculent brown precipitates. The $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$ molecules in the rhombohedral lattice have crystallographically imposed 32-site symmetry. As shown in Fig. 1, iron atoms define the apices of a twisted, penta-capped trigonal prism which is slightly irregular in the triclinic form. The oxo- and hydroxo-oxygen atoms are all triply bridging and form a distorted icosahedron with a close-packed, antiprismatic arrangement of the six μ_3 -oxo atoms at the core. The FeO_6 octahedra of the nine iron atoms comprising the trigonal prism and its three equatorial face caps are arranged in a non-planar ring with six μ_3 -hydroxo, corner-sharing oxygen atoms directed outwards (Fig. 2c). Two additional FeO_6 octahedra, situated on the 3-fold symmetry axis above and below the 9-membered ring, are linked to the aggregate by the 6 core μ_3 -oxo atoms. This type of condensation is believed to occur when $[\text{Fe}(\text{H}_2\text{O})_6]^{3+}$ ions polymerize with the

Table 1 Synthesis and crystallographic properties* of $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$

Synthesis		
$(\text{Et}_4\text{N})_2[\text{Fe}_2\text{OCl}_6] + 2 \text{Na}(\text{O}_2\text{CPh}) \xrightarrow{\text{CH}_3\text{CN}} \xrightarrow{\text{filter}} \text{red solution (57 mM in iron)}$		
$\xrightarrow{\text{precipitate}}$		
Solution	$\xrightarrow{\text{H}_2\text{O}}$	$[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}] \cdot 2\text{CH}_3\text{CN} \cdot n\text{H}_2\text{O}, n = 1-3$
	vapour diffusion (2 weeks)	
	0.8% H_2O in THF	
	liquid diffusion (1 month)	$[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}] \cdot 6\text{THF}, 57\% \text{ yield}$
X-ray structural data		
	$[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}] \cdot 2\text{CH}_3\text{CN} \cdot n\text{H}_2\text{O}$	$[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}] \cdot 6\text{THF}$
Unit cell	$a = 19.518(2) \text{ \AA}$	$a = 18.521(3) \text{ \AA}$
	$b = 22.696(3) \text{ \AA}$	$c = 46.659(6) \text{ \AA}$
	$c = 16.485(2) \text{ \AA}$	
	$\alpha = 94.60(1)^\circ$	
	$\beta = 93.47(1)^\circ$	
Space group	$\gamma = 81.61(1)^\circ$	
	$P\bar{1}$	$R\bar{3}2$ (hexagonal indexing)
Z	2	3
Current R -factor†	0.10	0.049

* Both structures were solved from single crystal X-ray diffractometer data by conventional direct methods and difference Fourier techniques followed by least-squares refinement of atomic positional and thermal parameters.

† $R = \sum ||F_o| - |F_c|| / \sum |F_o|$ where R is the crystallographic residual, and F_o and F_c are observed and calculated structure factors, respectively. Isotropic thermal parameters were used for all but the $\{\text{Fe}_{11}\text{O}_6(\text{OH})_6\}^{15+}$ core atoms, which were anisotropically refined, for the triclinic structure; anisotropic thermal parameters were used for all but the phenyl-ring carbon atoms of the rhombohedral structure.

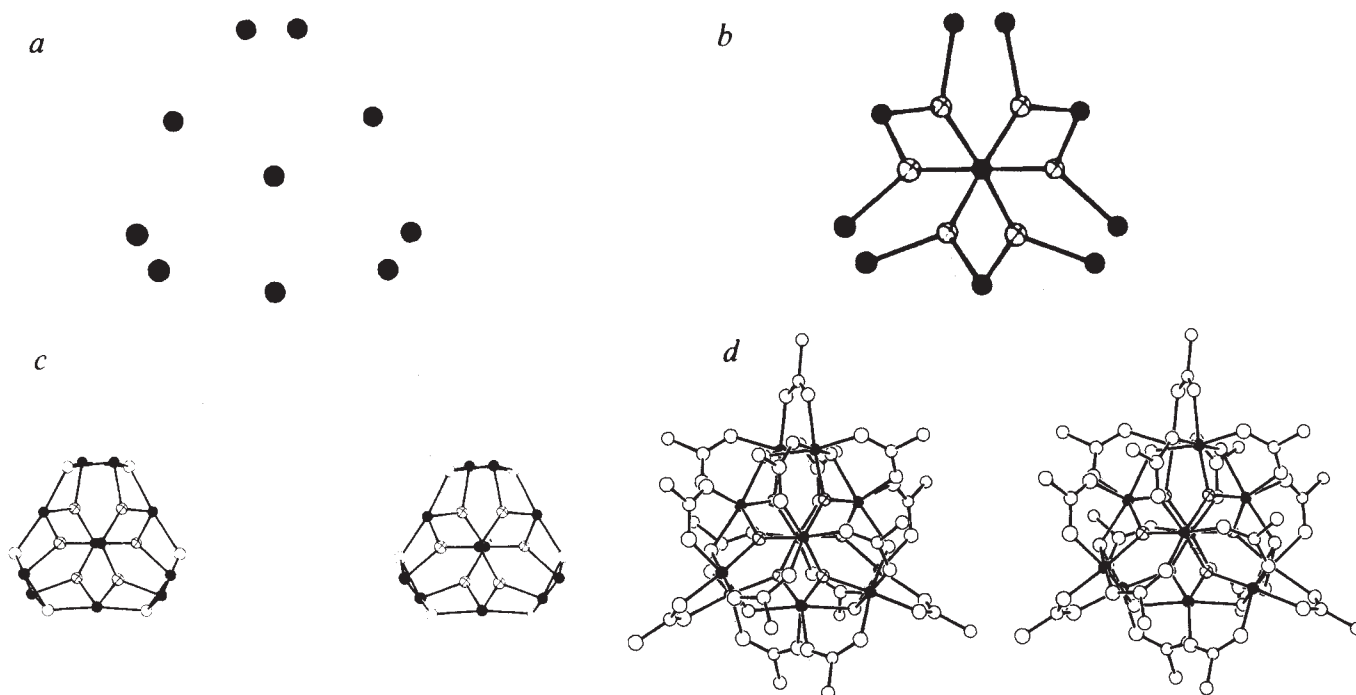


Fig. 2 Stepwise construction of $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$. *a*, Iron atoms viewed down the C_3 -axis; *b*, as *a* with addition of six μ_3 -oxo atoms; *c*, as *b* with addition of 6 μ_3 -hydroxo atoms (stereo view); *d*, as *c* with addition of 15 benzoate groups (stereo view). For clarity, only the first carbon atoms of the phenyl rings are depicted in *d*.

formation of μ -oxo and μ -hydroxo bridges⁴. The hydrophilic polymetallic core is encapsulated by a hydrophobic shell of 15 benzoate groups, but sufficient space is available for small molecules to approach the outwardly directed protons of the six hydroxo groups. Although we could not locate these hydroxyl hydrogen atoms crystallographically, their presence is supported by the molecular geometry (Fig. 1) and the fact that, in the rhombohedral lattice, six ordered THF molecules are hydrogen-bonded to the hydroxo groups with an $\text{O}(\text{H})\cdots\text{O}(\text{THF})$ distance of 2.78 Å. In addition, an $\text{O}-\text{H}$ stretching mode is observed in the infrared spectrum at $3,600\text{ cm}^{-1}$. A formal stepwise construction of the $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$ structure is given in Fig. 2.

The geometry (Fig. 1) and physical properties (Table 2) of $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$ are characteristic of high-spin ferric complexes with oxygen donor ligands. The magnetic moment is less than the spin-only value and decreases with decreasing temperature, consistent with overall anti-ferromagnetic interactions within the aggregate. Because the hydrophobic phenyl rings limit the size of the magnetic domain to that of the aggregate, no superparamagnetism analogous to that observed at low temperatures for ferritin¹¹ is observed down to 2.6 K (R. B. Frankel and G. C. Papaefthymiou, personal communication). Super-exchange within the aggregate is mediated by the bridging oxygen atoms in the usual manner⁹, although its details remain to be determined. There are three chemically distinct FeO_6 octahedra in the structure. From the north to the south poles along the C_3 axis, the 11 iron atoms are arranged in sets according to the scheme 1A, 3B, 3C, 3B, 1A, where A, B and C denote the three chemically distinct environments and the coefficients designate the number of iron atoms of each type. Consistent with this picture, preliminary Mössbauer spectra at 80 K, although unresolved, suggest the presence of inequivalent subsites (R. B. Frankel and G. C. Papaefthymiou, personal communication).

When iron(III) salts hydrolyse in aqueous solution, a mixture of structurally incompletely characterized polymers is obtained, the properties of which resemble those of the ferritin core^{5,12}. As these polymers age, a shoulder at 470 nm in the optical absorption spectrum increases in intensity, especially in the first

Table 2 Properties of $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$

Colour	Red-brown, $\lambda = 480\text{ nm}$ (shoulder)
Solubility	Soluble in acetonitrile, acetone and dimethylformamide
Infrared spectrum	$\nu_{\text{OH}} 3,600\text{ cm}^{-1}$
Magnetic moment	$4.2\ \mu_B$ per iron at 298 K
Mössbauer spectrum (zero field, 80 K)	Quadrupole splitting, 1.0 mm s^{-1} Isomer shift (versus iron metal at room temperature) 0.5 mm s^{-1}

12 h. This spectral change has been ascribed to formation of μ -oxo bridges from condensation of hydroxyl bridges with loss of water¹³. Acetonitrile solutions of $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$ are stable, however, as monitored by the constant appearance of a shoulder at 480 nm over the same time period. Thus, although protons on the hydroxyl groups are accessible for hydrogen bonding to small molecules such as water, THF or acetonitrile, the steric bulk of the phenyl rings protects the hydrophobic core of the undeca-iron(III) aggregate from further polymerization through condensation reactions.

In conclusion, the molecule $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$ demonstrates that a large polynuclear iron(III) complex can be assembled with biologically relevant ligands outside a protein environment. We achieved synthesis of the molecule through controlled hydrolytic polymerization of iron in aprotic media. The $[\text{Fe}_{11}\text{O}_6(\text{OH})_6]^{15+}$ core is anchored by carboxylate groups to phenyl rings which perform the function of the protein sheath by providing a hydrophobic environment formally analogous to that in ferritin. Although phosphate groups have not yet been introduced in this chemistry, they are known not to be an integral part of the iron-core structure of ferritin and their role in determining or stabilizing its crystalline structure is obscure¹⁴. Finally, the high 32-site symmetry of $[\text{Fe}_{11}\text{O}_6(\text{OH})_6(\text{O}_2\text{CPh})_{15}]$ is reminiscent of the 432-site symmetry of ferritin itself. Future work will test the properties of this molecule relevant to ferritin function, especially its redox behaviour, response to pH changes and metal removal in the presence of small chelating ligands

for iron, and will extend efforts to explore synthetically other possible polynuclear iron aggregates with biologically relevant ligands.

This work was supported by NIH grant GM-32134 from the National Institute of General Medical Sciences. Magnetic and Mössbauer measurements were made at the Francis Bitter National Magnet Laboratory at MIT, which is supported by NSF. We thank Drs R. B. Frankel and G. C. Papaefthymiou for preliminary Mössbauer measurements and helpful discussions.

Received 13 September; accepted 10 December 1985.

1. Mann, S., Frankel, R. B. & Blakemore, R. P. *Nature* **310**, 405-407 (1984).
2. Lowenstam, H. A. *Science* **211**, 1126-1130 (1981).
3. Ford, G. C. *et al. Phil. Trans. R. Soc. B304*, 551-565 (1984).
4. Chukrov, F. V., Zvyagin, B. B., Gorschkov, A. I., Yermilova, L. P. & Balashova, V. V. *Int. geol. Rev.* **16**, 1131-1143 (1974).
5. Spiro, T. G. *et al. J. Am. chem. Soc.* **88**, 2721-2726 (1966).
6. Wiegardt, K., Pohl, K., Jibril, I. & Huttner, G. *Angew. Chem. int. Ed. Engl.* **23**, 77-78 (1984).
7. Theil, E. C. *Adv. inorg. Biochem.* **5**, 1-38 (1983).
8. Chasteen, N. D., Antanaitis, B. C. & Aisen, P. J. *biol. Chem.* **260**, 2926-2929 (1985).
9. Gorun, S. M. & Lippard, S. J. *J. Am. chem. Soc.* **107**, 4568-4569 (1985).
10. Armstrong, W. H. & Lippard, S. J. *Inorg. Chem.* **24**, 981-982 (1985).
11. Blaise, A., Chappert, J. & Girardet, J. *C.R. hebdo. Séanc. Acad. Sci., Paris* **261**, 2310-2313 (1965).
12. Webb, J. & Gray, H. B. *Biochim. biophys. Acta* **351**, 224-229 (1974).
13. Sommer, B. A., Margerum, D. W., Renner, J., Saltman, P. & Spiro, T. G. *Bioinorg. Chem.* **2**, 295-309 (1973).
14. Macara, I. G., Hay, T. G. & Harrison, P. M. *Biochem. J.* **126**, 151-162 (1972).

Energetic cost of carrying loads: have African women discovered an economic way?

G. M. O. Maloiy*, N. C. Heglund†§, L. M. Prager†, G. A. Cavagna‡ & C. R. Taylor†

* Department of Animal Physiology, University of Nairobi, Nairobi, Kenya

† Harvard University, Concord Field Station, Bedford, Massachusetts 01730, USA

‡ University of Milan, Istituto di Fisiologia Umana, Milan 20133, Italy

When travelling in East Africa one is often surprised at the prodigious loads carried by the women of the area. It is not uncommon to see women of the Luo tribe carrying loads equivalent to 70% of their body mass balanced on the top of their heads (Fig. 1). Women of the Kikuyu tribe carry equally large loads supported by a strap across their foreheads; this frequently results in a permanently grooved skull. Recent experiments on running horses, humans, dogs and rats showed that the energy expended in carrying a load increased in direct proportion to the weight of the load for each animal at each speed, that is, carrying a load equal to 20% of body weight increased the rate of energy consumption by 20% (ref. 1). The purpose of the present study was to determine whether these African women use specialized mechanisms for carrying very large loads cheaply. We found that both the Luo and Kikuyu women could carry loads of up to 20% of their body weight without increasing their rate of energy consumption. For heavier loads there was a proportional increase in energy consumption, that is, a 30% load increased energy consumption by 10%, a 40% load by 20% and so on. We suggest that some element of training and/or anatomical change since childhood may allow these women to carry heavy loads economically.

Five African women walked on a motorized treadmill at five different speeds; each woman carried the loads in her customary manner. During this activity we measured the rate of oxygen consumption of the women by the open-flow method² in which



Fig. 1 Women of the Luo (left) and Kikuyu (right) tribes can often be seen carrying loads of wood, water or other provisions supported by their heads. These loads normally represent up to 70% of the woman's body weight. Drawing by Mr D. Njoroge.

air is drawn from a loose-fitting mask worn by the subject through a flow meter (Brooks variable-area float-type); the rate of flow of air through the mask is high enough ($123-142 \text{ l min}^{-1}$) to ensure that all of the expired air is collected. A small fraction of mask air is continuously analysed for oxygen content using a paramagnetic oxygen analyser (Taylor Servomex). The rate of oxygen consumption by the subject is calculated from the decrease in oxygen content of the air from the mask and the flow rate of air through the mask. Rate of oxygen consumption can be converted to rate of energy consumption because one litre of oxygen used in oxidative metabolism is approximately equal to 20.1 kJ. Stride frequency (F) was measured by timing 25 strides with a stopwatch.

The rate of oxygen consumption increased curvilinearly with speed during 'unloaded' walking in the African women (Fig. 2a). As a consequence of this curvilinear increase, the energy cost per unit distance is minimal at an intermediate, 'optimal' walking speed; this has been well documented previously for human walking^{3,4} (lower dotted line in Fig. 2a) and horse walking⁵, and is consistent with the currently understood mechanism of walking⁶. The 'optimal' speed corresponds to the speed at which the mechanical energy changes of the centre of mass are minimized by pendulum-like transfers between the kinetic and potential energy of the centre of mass. All of our measurements of oxygen consumption during load carrying were made at this walking speed.

When the African women walked with a 34-kg load (equivalent to 61% of the average body mass, M_b) their rate of oxygen consumption increased at all speeds by an average of 47% over the rate of oxygen consumption when they carried no load (Fig. 2a). However, this same load had no demonstrable effect on their energy consumption when they stood quietly (Fig.

Table 1 Proportional increase in oxygen consumption with respect to load when untrained adults carry loads on their backs or on their heads

	$\frac{\dot{V}_L/\dot{V}_U}{M_{\text{tot}}/M_b}$	F_L/F_U
Load on head	1.00 ± 0.04 (16)	1.00 ± 0.00 (15)
Load on back	1.00 ± 0.04 (22)	1.00 ± 0.00 (21)
Combined head and back	1.00 ± 0.04 (38)	1.00 ± 0.01 (36)

Values are mean $\pm$ s.d. Number of determinations shown in parentheses. M_{tot} , total mass; $\dot{V}$, rate of oxygen consumption. Stride frequency (F) is not changed by load carrying. Subscript U denotes determinations in which the subject was unloaded, L represents determinations in which a load was carried.

§ To whom correspondence should be addressed.

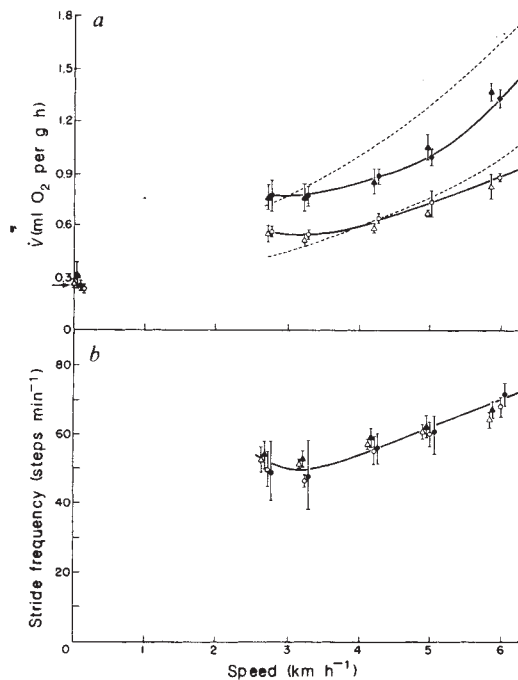


Fig. 2 *a*, Rate of energy consumption, measured as the rate of oxygen consumption ($\dot{V}$), was minimal at the 'optimal' walking speed of ~ 3 km h⁻¹ for both the unloaded Luo (Δ) and Kikuyu ($\circ$) women, and increased at lower and higher walking speeds. The vertical bars indicate the 95% confidence limits of the means. Walking with a 34-kg load increased the rate of energy consumption by an average of 47% in both the Luo (Δ) and Kikuyu ($\bullet$) women. There was no difference in the rate of oxygen consumption when the women stood quietly with or without the 34-kg load. For comparison, the lower dotted line indicates the oxygen consumption rate of 'unloaded' army recruits and the upper dotted line indicates the energetic cost of the recruits carrying 34-kg loads in their backpacks; the arrow on the ordinate indicates their metabolic rate when unloaded and motionless (army recruit data from Pandolf *et al.*⁴). *b*, Stride frequency did not change when the women either carried a 34-kg load or had no load. Symbols same as for *a*.

2*a*). Stride frequency was not altered at any speed by carrying a load (Fig. 2*b*).

Our most interesting finding was that both the Luo and Kikuyu women carried loads of up to 20% of their body weight without energetic cost (Fig. 3). Increasing the loads from 20% to 70% of body weight increased the energetic cost linearly from 0% to 50% (solid line in Fig. 3). Thus, the relationship between load and energy consumption seen in previous experiments on running animals¹ was simply offset by the initial 'free' 20% of M_b load in the walking African women (Fig. 3).

Is supporting the weight on the head a particularly economic way of carrying loads? Pandolf *et al.* found that male army recruits carrying a 34-kg load (61% of M_b) in a backpack would have a higher rate of oxygen consumption than the African women carrying similar loads on their heads at all except the very lowest walking speeds⁴ (upper dotted line in Fig. 2*a*). At the 'optimal' walking speed a load equivalent to 70% of M_b would increase the recruits' oxygen consumption rate nearly twofold compared with the unloaded rate, but only 1.5–1.6-fold in the African women. In addition, the army recruits were not able to carry small loads without metabolic cost; their oxygen consumption rate increased by 13% when carrying a load of 20% of their body weight in a backpack.

Is it possible that people not trained from childhood can also carry loads on their head relatively cheaply? We measured the oxygen consumption rate and stride frequency in three adults (two males and one female) walking on a treadmill with (1) no load, (2) a load on the head, and (3) a load in a backpack. We

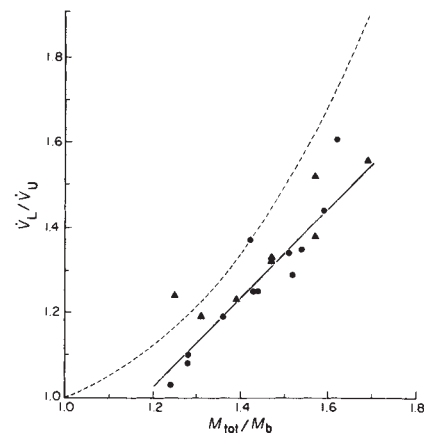


Fig. 3 Ratio of the rate of oxygen consumption ($\dot{V}$) during loaded walking (L) to that during unloaded walking (U) plotted against the ratio of the total mass (M_b + load) to the unloaded mass (M_b) for the Luo (Δ) and Kikuyu ($\bullet$) women walking at the 'optimal' speed of 3.25 km h⁻¹. The rate of oxygen consumption was unchanged (that is, the oxygen consumption ratio was 1.0) for loads of up to 20% of the individual woman's mass (that is, at a mass ratio of 1.2). At higher loads the increase in rate of oxygen consumption was proportional to the increase in load. This relation is parallel to the direct proportion previously found during load-carrying in running animals¹ and found in the present study for head or backpack load-carrying in untrained individuals, but is decreased by a constant amount due to the 'free' load of up to 20% M_b . The broken line indicates the expected relation for army recruits, as calculated from Pandolf *et al.*⁴.

found that the ratio of the increase in oxygen consumption to the increase in weight was unity for both the loads carried on the back and the loads carried on the head (Table 1). It was more difficult for these untrained individuals to carry loads on their heads than on their backs; they could safely manage loads of up to 40% M_b on their backs, but loads of only up to 15% M_b on their heads. Stride frequency was unchanged with loading.

How are the African women able to carry such large loads on their heads so economically? It seems reasonable to assume that the energy requirements for moving the legs and arms relative to the centre of mass would be the same for loaded and unloaded people because stride frequency is unchanged when carrying a load. However, the energy requirements for lifting the centre of mass within each step, and for accelerating and decelerating the centre of mass at each step, would be proportional to the total mass (M_b + load) if the movements of the centre of mass were unchanged; the African women may be more economical simply because they are able to minimize the movements of the load when supporting the load with their heads (the available experimental evidence shows that the amplitude of the vertical movements during normal unloaded walking decreases as one moves up the spine from the coccyx to the head⁷). The fact that the African women alone can carry loads of up to 20% M_b without any measurable metabolic cost may indicate that some anatomical change has occurred, as a result of carrying large loads since childhood, which allows these women to support small loads using non-metabolizing structural elements.

This study was supported by the NIH, NSF, National Geographic Society, the Guggenheim Foundation, the Nuffield Foundation, Wellcome Trust and the Leverhulme Trust.

Received 18 November 1985; accepted 2 January 1986.

1. Taylor, C. R., Heglund, N. C., McMahon, T. A. & Looney, T. R. *J. exp. Biol.* **86**, 9–18 (1980).
2. Fedak, M. A., Rome, L. & Seherman, H. J. *J. appl. Physiol.* **51**(3), 772–776 (1981).
3. Margaria, R. *Biomechanics and Energetics of Muscular Exercise* (Clarendon, Oxford, 1976).
4. Pandolf, K. B., Givoni, B. & Goldman, R. F. *J. appl. Physiol.* **43**(4), 577–581 (1977).
5. Hoyt, D. F. & Taylor, C. R. *Nature* **292**, 239–240 (1981).
6. Cavagna, G. A., Thys, H. & Zamboni, A. *J. Physiol., Lond.* **262**, 639–657 (1976).
7. Cavagna, G. A., Saibene, F. P. & Margaria, R. *J. appl. Physiol.* **19**(2), 249–256 (1964).

GTP-binding proteins mediate transmitter inhibition of voltage-dependent calcium channels

George G. Holz IV, Stanley G. Rane
& Kathleen Dunlap

Department of Physiology, Tufts University School of Medicine,
Boston, Massachusetts 02111, USA

The modulation of voltage-dependent calcium channels by hormones and neurotransmitters has important implications for the control of many Ca^{2+} -dependent cellular functions including exocytosis and contractility¹⁻⁷. We made use of electrophysiological techniques, including whole-cell patch-clamp recordings from dorsal root ganglion (DRG) neurones, to demonstrate a role for GTP-binding proteins (G-proteins) as signal transducers in the noradrenaline- and γ -aminobutyric acid (GABA)-induced inhibition of voltage-dependent calcium channels⁸⁻¹¹. This action of the transmitters was blocked by: (1) preincubation of the cells with pertussis toxin (a bacterial exotoxin catalysing ADP-ribosylation of G-proteins¹²); or (2) intracellular administration of guanosine 5'-O-(2-thiodiphosphate) (GDP- β -S), a non-hydrolysable analogue of GDP that competitively inhibits the binding of GTP to G-proteins¹³. Our findings provide the first direct demonstration of the G-protein-mediated inhibition of voltage-dependent calcium channels by neurotransmitters. This mode of transmitter action may explain the ability of noradrenaline and GABA to presynaptically inhibit Ca^{2+} -dependent neurosecretion from DRG sensory neurones^{4,5}.

Recordings were obtained from primary cultures of embryonic chick DRG cells (see Fig. 1). When bathed in solutions containing 1 mM Ba^{2+} and 2 mM Ca^{2+} , these cells generate action potentials with a prominent calcium-dependent plateau phase. The plateau results from a regenerative inward current carried by Ba^{2+} and Ca^{2+} ions and is blocked by cobalt¹⁴. Previous studies have demonstrated that noradrenaline and GABA decrease the duration of these action potentials by inhibiting the depolarization-induced calcium current^{8,9}. In the present study, a saturating concentration of noradrenaline (50 μM) reduced the action-potential duration in 75% of the cells tested by an average of $43 \pm 4.7\%$ (Fig. 1a, Table 1). The inhibitory action of noradrenaline was also observed as a decrease in the Ca current recorded from voltage-clamped DRG cells (Fig. 1b). Similar decreases in action-potential duration and Ca current were also observed during application of GABA (Table 1).

Pertussis toxin (PTX) blocks G-protein-mediated responses to hormones and neurotransmitters in many cell types¹². Specifically, PTX catalyses ADP-ribosylation of G-proteins, thereby preventing agonist-induced dissociation of the proteins into active subunits^{15,16}. When applied to DRG cells, PTX inhibits the transmitter-induced decrease in action-potential duration (Table 1). Following exposure to PTX (140 ng ml^{-1}), only 9 and 19% of the cells responded to noradrenaline and GABA, respectively. Note that the mean percentage decrease in action-potential duration for those PTX-treated cells which did respond to noradrenaline and GABA was also reduced relative to control. The responses recorded from PTX-treated cells were very slow in onset: the maximal decrease in action-potential duration was observed only after continued application of noradrenaline or GABA for 1-2 min. In contrast, noradrenaline and GABA responses recorded from cells untreated with PTX were rapid in onset, reaching a maximal decrease in action-potential duration after only 20-30 s of continual application.

Recordings from voltage-clamped DRG cells demonstrated that, as expected, PTX also blocked the transmitter-induced decrease in Ca current. Before PTX treatment, noradrenaline

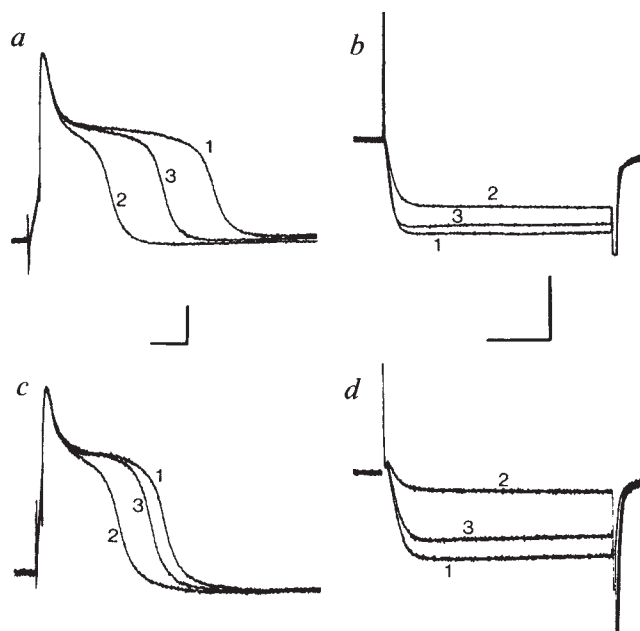


Fig. 1 Noradrenaline and OAG decrease action-potential duration and Ca current in chick DRG cells. *a, c*, Current-clamp recordings of DRG cell action potentials; *b, d*, voltage-clamp recordings of Ca currents. Traces were recorded before (1), during (2) and after (3) treatment with either 50 μM noradrenaline (*a, b*) or 60 μM OAG (*c, d*). Action potentials were evoked by direct current injection through the recording electrode. Ca currents were evoked by a +50 mV step depolarization from a holding potential of -60 mV. At this holding potential the predominant inward current is through non-inactivating Ca channels. Scale bars: *a*, 20 mV, 5 ms; *b*, 4 nA, 40 ms; *c*, 20 mV, 10 ms; *d*, 2 nA, 40 ms.

Methods. Primary cultures of chick DRG cells were prepared as previously described^{8,9}. Briefly, ganglia from 11-12 day-old embryos were dissociated in a Ca^{2+} - and Mg^{2+} -free Pucks solution, suspended in MEM (supplemented with nerve growth factor, 5% chick embryo extract, 10% horse serum, 2 mM glutamine, penicillin (50 U ml^{-1}) and streptomycin (50 $\mu\text{g ml}^{-1}$), γ -irradiated (5,000 rads) and plated in 35-mm collagen-coated tissue culture dishes. Action potentials were recorded using an amplifier with an active bridge circuit allowing current injection through the recording microelectrodes (40-80 M Ω , filled with 2 M KCl). The bathing solution contained (mM): 132 NaCl, 2.5 KCl, 2.0 CaCl_2 , 1.0 BaCl_2 , 0.8 MgCl_2 and 25 HEPES (pH 7.4). Ca currents were recorded as described previously¹⁷, using the whole-cell configuration of the patch-clamp technique¹⁸. The pipette solution contained (mM): 150 CsCl, 5 bis(*o*-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid (BAPTA), 5 MgATP and 10 HEPES (pH 7.3); the external bathing solution contained 133 NaCl, 1 CaCl_2 , 10 tetraethylammonium, 0.3 μM tetrodotoxin and 25 HEPES (pH 7.3). Little or no rundown of the Ca current was observed during the first 10-15 min of recording. Noradrenaline, GABA and OAG (Sigma) were dissolved in the appropriate bathing solution and pressure ejected from blunt-tipped pipettes as previously described⁹.

(10 μM) reduced the Ca current by $35 \pm 6.0\%$ in seven of 8 cells tested. After treatment with 140 ng ml^{-1} PTX, the fraction of cells responding to noradrenaline was reduced to two of eight, and the average decrease in Ca current was only $13 \pm 2.4\%$. The action of PTX seems to be selective for receptor-mediated alterations in Ca channel function: PTX did not block responses to the diacylglycerol analogue 1,2-oleoyl acetylgllycerol (OAG), an activator of protein kinase C that mimics the effects of noradrenaline and GABA on DRG cells¹⁷. A saturating concentration of OAG (60 μM) decreased the action-potential duration by an average of $37 \pm 2.9\%$ in 74% of the cells tested ($n=19$) and decreased the Ca current by $38 \pm 5.3\%$ in five of five cells tested (Fig. 1c, d). In cultures treated with

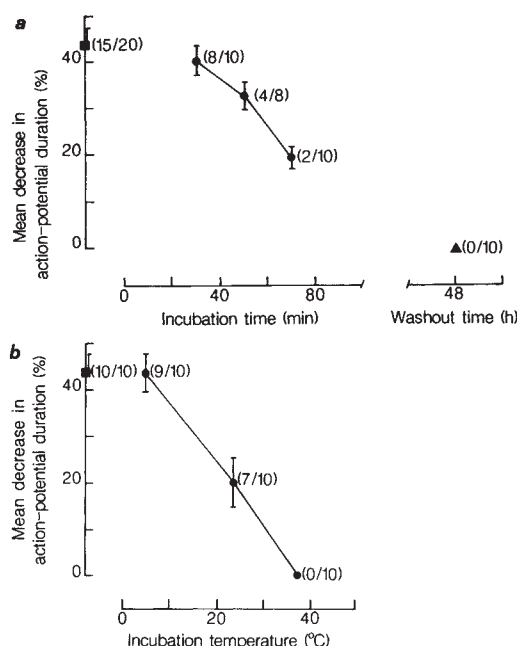


Fig. 2 Time- and temperature-dependent blockade by PTX of DRG cell responses to noradrenaline. **a**, Time dependence of the action of PTX. DRG cell cultures were incubated in MEM containing 140 ng ml^{-1} PTX at 37°C for 30, 50 and 70 min (circles). Control cultures (square) were incubated in MEM (without added PTX or $(\text{NH}_4)_2\text{SO}_4$) at 37°C for 1 h. Recovery from PTX-treatment was tested at 48 h following washout of PTX (triangle). The results are expressed as the mean ($\pm$ s.e.m.) percentage decrease in action-potential duration in response to $50 \mu\text{M}$ noradrenaline. In parentheses is indicated the fraction of cells tested that responded to noradrenaline. **b**, Temperature dependence of the action of PTX. Cultures were incubated in MEM containing 140 ng ml^{-1} PTX for 4 h at 5, 23 and 37°C (circles). Control cultures were incubated in MEM for 4 h at 5°C (square). The fraction of cells responding to noradrenaline is also indicated in parentheses. The medium was HEPES-buffered and equilibrated with air when incubating at 5 and 23°C . All results in **a** and **b** were obtained from cultures of the same plating.

PTX (140 ng ml^{-1}), the responses to OAG were not attenuated.

ADP-ribosylation of G-proteins by PTX requires prior internalization and activation of the toxin¹². The effects of PTX, therefore, occur with some delay. When PTX was tested on DRG cells, we found its action to be slow in onset and prolonged in duration (Fig. 2a). A progressive decrease in the number of cells responding to noradrenaline was observed after treatment with PTX for 30, 50 and 70 min. A progressive decrease in the magnitude of the response to noradrenaline was also observed. A lag time of ~ 30 min preceded the inhibitory actions of PTX. To test for recovery, cultures were exposed to 140 ng ml^{-1} PTX for 4 h, washed repeatedly with minimal essential medium (MEM) and re-incubated in culture medium. Even after 48 h of recovery, a total blockade of transmitter action was observed (Fig. 2a).

As expected for a process of internalization, the action of PTX was temperature-dependent. Incubation of cultures in MEM containing 140 ng ml^{-1} PTX for 4 h at 5°C did not diminish the response to noradrenaline (Fig. 2b). When cultures were incubated in PTX for 4 h at 23°C , the number of cells responding to noradrenaline was reduced, and the magnitude of the response decreased relative to the control response. With an increase in incubation temperature to 37°C , a total blockade of the action of noradrenaline was observed.

The blockade of noradrenaline and GABA responses by PTX suggests that G-proteins mediate the inhibitory actions of trans-

Table 1 Pertussis toxin blocks DRG cell responses to noradrenaline and GABA

	Noradrenaline		GABA	
	Cells responding	Mean decrease in action-potential duration (%)	Cells responding	Mean decrease in action-potential duration (%)
Control	15/20	43 ± 4.7	16/19	44 ± 4.6
PTX-treated	2/22	24 ± 5.0	3/16	36 ± 2.5
Vehicle-treated	9/10	45 ± 6.0	—	—

DRG cell responses to $50 \mu\text{M}$ noradrenaline and $50 \mu\text{M}$ GABA were recorded as a decrease in action-potential duration measured at 1/2 peak spike amplitude. Only reversible and repeatable responses were included in the results. Under the recording conditions used, measurements of decreases in action-potential duration of $<10\%$ were considered unreliable and were therefore scored as no effect. PTX was stored at 4°C as a stock suspension (1.4 mg ml^{-1}) in saturated $(\text{NH}_4)_2\text{SO}_4$. The stock suspension was diluted 1:1,000 in 10 mM sodium phosphate buffer ($\text{pH } 7.2$) containing 50 mM NaCl and 0.04% heat-inactivated bovine serum albumin. Freshly diluted PTX was then diluted 10-fold in MEM containing 0.1% glutamine (no horse serum, nerve growth factor or penicillin-streptomycin added) to give a final dilution factor of 1:10,000 containing 140 ng ml^{-1} PTX. DRG cell cultures were incubated in 2 ml of 140 ng ml^{-1} PTX at 37°C for 4–8 h. Control cultures were incubated at 37°C for 4–8 h in MEM containing sodium phosphate buffer but with no added PTX or $(\text{NH}_4)_2\text{SO}_4$. Vehicle-treated cultures were incubated at 37°C for 4–8 h in MEM containing saturated $(\text{NH}_4)_2\text{SO}_4$ diluted 1:10,000. All results were obtained from cultures of the same plating. Results are expressed as mean $\pm$ s.e.m. We did not observe any direct effects of PTX on the electrical properties of the neurones. The mean resting membrane potential, action-potential duration and action-potential amplitude of PTX-treated cells did not differ significantly from that of untreated cells. Furthermore, the mean amplitude of the Ca current recorded from voltage-clamped cells was similar for both PTX-treated and untreated cells.

mitters on neuronal Ca channels. To substantiate such a role for G-proteins, DRG cells were voltage-clamped and loaded with GDP- β -S by the whole-cell recording variation of the patch-clamp technique¹⁸. GDP- β -S competes with GTP for the guanine nucleotide binding site on G-proteins, thereby blocking GTP-dependent activation of the proteins by hormones and neurotransmitters^{19–21}. Intracellular dialysis of DRG cells with GDP- β -S (100 – $500 \mu\text{M}$) blocked the noradrenaline-induced decrease in Ca current in a dose-dependent manner (Fig. 3). The blockade of noradrenaline responses was observed using concentrations of GDP- β -S similar to that reported to block adrenergic receptor-mediated inhibition of adenylate cyclase in human platelets²⁰. Additional experiments demonstrated that GDP- β -S does not interfere with the action of OAG on DRG cells. Before exposure to GDP- β -S, OAG reduced the Ca current by $38 \pm 5.3\%$ ($n = 5$), whereas after treatment with $250 \mu\text{M}$ GDP- β -S, OAG decreased the Ca current by $35 \pm 3.0\%$ ($n = 5$).

The noradrenaline and GABA receptors mediating inhibition of DRG cell Ca channels are similar to α_2 adrenergic²² and GABA-B²³ receptors, two receptor subtypes known to be negatively coupled to adenylate cyclase^{24,25}. The actions of GDP- β -S and PTX on chick DRG cells may therefore result from their ability to block noradrenaline and GABA receptor-mediated activation of the G-protein (N_i) promoting transmitter inhibition of adenylate cyclase^{15,20,26,27}. In this manner, noradrenaline and GABA would inhibit Ca channel function by lowering intracellular concentrations of cyclic AMP. To test this hypothesis, we dialysed DRG cells with solutions containing 5 mM cAMP and $250 \mu\text{M}$ 3-isobutyl-1-methylxanthine (a phosphodiesterase inhibitor). As previously reported²⁸, cAMP did not attenuate the noradrenaline-induced inhibition of the Ca current ($N = 5$

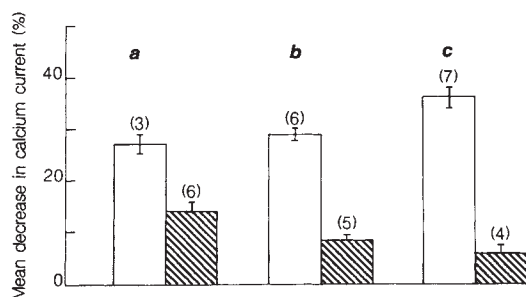


Fig. 3 Blockade by GDP-β-S of DRG cell responses to noradrenaline. DRG cells were dialysed for 5 min with 100 (a), 250 (b) and 500 (c) μM GDP-β-S (trilithium salt, Boehringer Mannheim) by inclusion of the GDP analogue in the patch pipette solution, which allows diffusional exchange between the intracellular fluid contents and the pipette solution. The Ca current was evoked from a holding potential of -60 mV in response to depolarizing test pulses of 50–60 mV. The amplitude of the Ca current was monitored at 10-s intervals. The mean decrease in Ca current in response to noradrenaline (10 μM) is indicated for untreated control cells (unshaded columns) and cells dialysed with GDP-β-S (shaded columns). The responses to noradrenaline obtained from cells treated with GDP-β-S were compared with responses recorded from untreated cells in the same culture dish. All results shown were obtained from cultures of the same plating. Error bars indicate s.e.m. (number of cells is indicated in parentheses). GDP-β-S did not directly inhibit the Ca current, nor did it accelerate run-down of the current: the mean amplitude of the current recorded from 250 μM GDP-β-S-treated and untreated cells was similar after dialysis for 5 min (treated, 2.7 ± 0.3 nA, $n = 10$; untreated, 3.0 ± 0.4 nA, $n = 11$). Dialysis of DRG cells with 1.5 mM LiCl ($n = 9$) did not reduce the response to noradrenaline relative to the control response.

cells). These findings indicate that in DRG cells the PTX substrate mediating transmitter inhibition is a G-protein structurally related to N_i , but not directly coupled to adenylate cyclase. One possibility is that this DRG cell G-protein corresponds to the PTX-sensitive N_0 α-protein of relative molecular mass 39,000 isolated from bovine brain^{29,30}.

Previous studies support a role for G-proteins in the receptor-mediated activation of membrane phospholipases^{31–35}. It remains to be determined whether a similar G-protein-mediated stimulation of phospholipase activity underlies the inhibitory actions of noradrenaline and GABA on neuronal Ca channels. The best evidence implicating phospholipases in the inhibition of neuronal Ca channels is the demonstration that the protein kinase C activator OAG³⁶ blocks the DRG cell Ca current in a manner similar to that of noradrenaline and GABA¹⁷. Alternatively, G-proteins may directly couple transmitter receptors to the Ca channel. Although these points remain to be resolved, our findings clearly indicate that cellular processes controlling Ca homeostasis are influenced by alterations in the structure and function of G-proteins. In terms of neuronal function, therefore, G-proteins are likely to serve as important intermediaries in processes governing receptor-mediated inhibition or facilitation of Ca^{2+} -dependent neurosecretion.

We thank Dr Ron Sekura for the gift of PTX and Drs Paul Brehm and Michael Goy for their advice. This work was supported by grants to K.D. from the Klingenstein Foundation and the American Heart Association.

Received 20 September; accepted 20 December 1985.

- Kupferman, I. A. *Rev. Neurosci.* **2**, 447–465 (1979).
- Shain, W. & Carpenter, D. O. *Int. Rev. Neurobiol.* **22**, 205–250 (1981).
- Dunlap, K. in *The Mechanism of Gated Calcium Transport across Biological Membranes* (eds Ohnishi, S. T. & Endo, M.) 87–97 (Academic, New York, 1981).
- Fischbach, G. D., Dunlap, K., Mudge, A. W. & Leeman, S. in *Neurosecretion and Brain Peptides* (eds Martin, J. B., Reichlin, S. & Bick, K. L.) 175–188 (Raven, New York, 1981).
- Holz, G. G., Kream, R. M. & Dunlap, K. *Soc. Neurosci. Abstr.* **11**, 126 (1985).
- Holz, G. G., Shefner, S. A. & Anderson, E. *G. J. Neurosci.* (in the press).
- Reuter, H. & Scholz, H. *J. Physiol., Lond.* **264**, 49–62 (1977).

- Dunlap, K. & Fischbach, G. *J. Physiol., Lond.* **317**, 519–535 (1981).
- Dunlap, K. & Fischbach, G. *Nature* **276**, 837–839 (1978).
- Houslay, M. D. *Trends biochem. Sci.* **9**, 39–40 (1984).
- Gilman, A. G. *Cell* **36**, 577–579 (1984).
- Ui, M. *Trends pharmac. Sci.* **5**, 277–279 (1984).
- Eckstein, F., Cassel, D., Levkovitz, H., Lowe, M. & Selinger, Z. *J. Biol. Chem.* **254**, 9829–9834 (1979).
- Dichter, M. & Fischbach, G. D. *J. Physiol., Lond.* **267**, 281–298 (1977).
- Katada, T., Bokoch, G. M., Northup, J. K., Ui, M. & Gilman, A. G. *J. Biol. Chem.* **259**, 3568–3577 (1984).
- Jakobs, K. H., Aktories, K. & Schultz, G. *Adv. Cyclic Nucleotide Res. Protein Phosphor.* **17**, 135–143 (1984).
- Rane, S. G. & Dunlap, K. *Proc. natn. Acad. Sci. U.S.A.* **83**, 184–188 (1986).
- Hamill, O., Marty, A., Neher, E., Sakman, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
- Cassel, D., Eckstein, F., Lowe, M. & Selinger, Z. *J. Biol. Chem.* **254**, 9835–9838 (1979).
- Jakobs, K. H. *Eur. J. Biochem.* **132**, 125–130 (1983).
- Lemos, J. R. & Levitan, I. B. *J. gen. Physiol.* **83**, 269–285 (1984).
- Canfield, D. R. & Dunlap, K. *Br. J. Pharmac.* **82**, 557–563 (1984).
- Dunlap, K. *Br. J. Pharmac.* **74**, 579–585 (1981).
- Sabol, S. L. & Nirenberg, M. *J. Biol. Chem.* **254**, 1913–1920 (1979).
- Wojcik, W. J. & Neff, N. H. *Molec. Pharmac.* **25**, 24–28 (1984).
- Hazeki, O. & Ui, M. *J. Biol. Chem.* **256**, 2856–2862 (1981).
- Cote, T. E., Frey, E. A. & Sekura, R. D. *J. Biol. Chem.* **259**, 8693–8698 (1984).
- Forscher, P. & Oxford, G. S. *J. gen. Physiol.* **85**, 743–763 (1985).
- Neer, E., Lok, J. & Wolf, L. *J. Biol. Chem.* **259**, 14222–14229 (1984).
- Sternweis, P. C. & Robishaw, J. D. *J. Biol. Chem.* **259**, 13806–13813 (1984).
- Nakamura, T. & Ui, M. *J. Biol. Chem.* **260**, 3584–3593 (1985).
- Bradford, P. G. & Rubin, R. P. *FEBS Lett.* **183**, 317–320 (1985).
- Volpi, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2708–2712 (1985).
- Cockcroft, S. & Comperts, B. *Nature* **314**, 534–536 (1985).
- Wallace, M. A. & Fain, J. N. *J. Biol. Chem.* **260**, 9527–9530 (1985).
- Nishizuka, Y. *Nature* **308**, 693–698 (1984).

Identity of cells that imprint H-2-restricted T-cell specificity in the thymus

David Lo & Jonathan Sprent

Department of Immunology, IMM4, Scripps Clinic and Research Foundation, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

The thymus has two important roles in controlling the specificity of T lymphocytes. First, T cells differentiating in the thymus are rendered tolerant of 'self' antigens, particularly antigens encoded by the major histocompatibility complex, the H-2 complex in mice¹. Second, the thymus imbues T cells with the property of H-2-restricted recognition of antigen, that is, the capacity of T cells to react with foreign antigens presented in association with self H-2 gene products^{2,3}. Until recently it has generally been assumed that self-tolerance and H-2-restricted specificity both reflect early T-cell contact with self H-2 determinants expressed on thymic epithelial cells. Recent evidence suggests, however, that intrathymic cells of the macrophage/dendritic cell (MØ/DC) lineage also have a role in shaping T-cell specificity^{4–7}. In particular, it has been found that the tolerance to graft-type H-2 determinants which normally ensues when T cells differentiate in an H-2-different thymus fails to occur when the thymus is pre-treated with deoxyguanosine (dGuo)^{6,7}, a procedure that selectively destroys MØ/DC but spares epithelial cells⁸. In contrast to these findings on tolerance induction, evidence is presented here that dGuo-treated thymus grafts do imprint T cells with H-2-restricted specificity for antigen. It appears, therefore, that induction of tolerance and H-2 restriction are controlled by different cells in the thymus.

The main aim of the experiments described here was to determine whether intrathymic MØ/DC have a role in imprinting T cells with H-2-restricted specificity. Evidence that the thymus controls H-2 restriction has come largely from studies with bone marrow chimaeras and thymus-grafted mice. The key finding from such experiments is that H-2-heterozygous ($a \times b$)F₁ hybrid stem cells differentiating in a strain a thymus respond preferentially to antigen presented by strain a (thymic) cells rather than strain b (non-thymic) cells. This preference is especially pronounced in the case of helper T (T_h) cells, that

Table 1 H-2-restricted specificity of T cells from bone marrow chimaeras and thymus-grafted mice

Donors of LN T cells used in proliferation assay	Mouse no.	³ H-TdR incorporation (c.p.m. × 10 ³) by LN T cells cultured with irradiated spleen cells				KLH response to B10.BR (Δ c.p.m.)/ KLH response to B6 (Δ c.p.m.)
		B6	B6 + KLH	B10.BR	B10.BR + KLH	
<i>a</i> , F ₁ → F ₁ chimaeras	1	0.4	25.2	0.2	13.3	0.53
	2	0.4	19.9	0.3	15.1	0.76
	3	0.2	11.4	0.2	8.1	0.71
						Mean = 0.67
<i>b</i> , F ₁ → B6 chimaeras	1	0.4	24.9	0.5	0.8	0.01
	2	0.6	37.3	0.3	1.6	0.04
	3	0.4	43.6	0.3	0.8	0.01
	4	0.7	47.5	0.3	2.4	0.04
						Mean = 0.03
<i>c</i> , ATx F ₁ given B6 fetal thymus 2 weeks before irradiation (1,100 rad)	1	0.4	21.1	0.3	0.7	0.02
	3	0.2	23.9	0.3	1.1	0.04
	4	0.6	47.6	0.5	2.9	0.05
	5	0.4	47.0	0.2	1.4	0.03
						Mean = 0.04
<i>d</i> , ATx F ₁ given B6 thymus 2 months before 1,100 rad	1	0.5	48.1	0.3	2.8	0.05
	2	0.6	61.3	0.4	6.0	0.09
	3	0.3	32.7	0.5	1.5	0.03
						Mean = 0.06
<i>e</i> , ATx F ₁ given dGuo-treated B6 fetal thymus	1	0.5	26.0	0.3	5.1	0.18
	2	0.2	31.0	0.2	1.5	0.04
	3	0.3	25.0	0.3	4.4	0.17
	4	0.4	24.2	0.3	0.7	0.02
	5	0.8	35.2	0.3	2.6	0.07
						Mean = 0.10
<i>f</i> , F ₁ → CBA chimaeras	1	0.2	0.4	0.4	28.1	112
	2	0.1	0.5	0.5	27.0	70
	3	0.4	0.7	0.6	27.0	80
	4	0.1	2.1	0.3	40.5	20
						Mean = 71
<i>g</i> , ATx F ₁ given CBA fetal thymus	1	0.1	0.4	0.2	30.4	89
	2	0.2	0.3	0.4	22.1	148
	3	0.2	3.4	0.4	41.1	13
	4	0.4	1.2	0.9	30.7	44
	5	0.2	0.6	0.3	33.3	70
						Mean = 73
<i>h</i> , ATx F ₁ given dGuo-treated CBA fetal thymus	1	0.1	1.6	0.1	31.3	21
	2	0.3	2.0	0.7	54.8	31
	3	0.7	0.9	0.5	74.8	464
						Mean = 172

Bone marrow chimaeras (groups *a*, *b*, *f*) were prepared by exposing 8–12-week-old B6, CBA or (B6 × CBA)F₁ mice to 1,100 rad and then reconstituting the mice with T-cell-depleted (anti-Thy 1.2 + complement-treated) (B6 × CBA)F₁ bone marrow cells given intravenously. For thymus-grafted mice (groups *c*, *d*, *e*, *g*, *h*), (B6 × CBA)F₁ mice (6 weeks old) were thymectomized 2–6 weeks before thymus grafting. For grafting, day-14 fetal thymuses were cultured overnight in clusters of 5–6 lobes to allow consolidation and then placed under the kidney capsule; thymuses were pretreated *in vitro* with dGuo as described in Fig. 1 legend. Thymus-grafted mice were exposed to 1,100 rad and reconstituted with anti-Thy 1.2 + complement-treated (B6 × CBA)F₁ bone marrow⁹ 2 weeks after grafting, except in group *d*, where animals were irradiated 2 months after grafting. Two months after irradiation and marrow reconstitution, all mice were primed in the rear footpads and tailbase with 20–30 μg KLH (Sigma) in complete Freund's adjuvant; 9 days after priming, cells from draining lymph nodes (LN) were passed over nylon wool columns¹⁸, treated with a mixture of anti-I-A^k and anti-I-A^b monoclonal antibodies plus complement, and placed in culture (10⁵ per well) with 2 × 10⁵ stimulator cells (spleen cells pulsed for 2–4 h with 100 μg ml⁻¹ KLH), washed, and irradiated (1,500 rad). Cultures were pulsed with 1 μCi ³H-TdR on day 3 and collected on day 4. Data are mean of triplicate cultures; s.d. (omitted for simplicity) was generally within 10–20% of the mean.

is, cells that are restricted by class II (I-A, I-E) (Ia) H-2 determinants^{9–11}. With cytotoxic T (T_c) cells—cells restricted by class I (H-2K and H-2D) H-2 determinants—strong restriction to thymic H-2 determinants has been found by some groups^{12–14}, but not others^{15–17}.

In the case of T_h cells, T cells in F₁ → parent chimaeras remain heavily restricted to thymic (host) H-2 determinants for a year or more after reconstitution⁹. At face value, this finding would seem to argue against a role for intrathymic MØ/DC in imprinting H-2 restriction, since donor cells of this type would be expected to enter the thymus rapidly. According to Longo and Schwartz⁴, however, these cells take up to 2 months to enter the thymus. From a variety of experiments, including studies in

which long-term F₁ → parent chimaeras were depleted of T cells and then allowed to generate a second set of T cells, these workers argue that only MØ/DC and not epithelial cells imprint intrathymic H-2 restriction⁴. A corollary to this viewpoint is that T cells differentiating in thymus grafts depleted of MØ/DC would not show restriction to thymic H-2 determinants.

The results in Table 1 show secondary proliferative responses of T cells from F₁ → parent chimaeras or thymus-grafted mice primed to keyhole limpet haemocyanin (KLH) in adjuvant¹⁸. F₁ → parent chimaeras were prepared by exposing adult CBA/Ca (CBA) (H-2^k) or C57BL/6 (B6) (H-2^b) mice to heavy irradiation (1,100 rad) and then reconstituting the mice with T-cell-depleted (B6 × CBA)F₁ marrow (see Table 1 legend). Thymus-

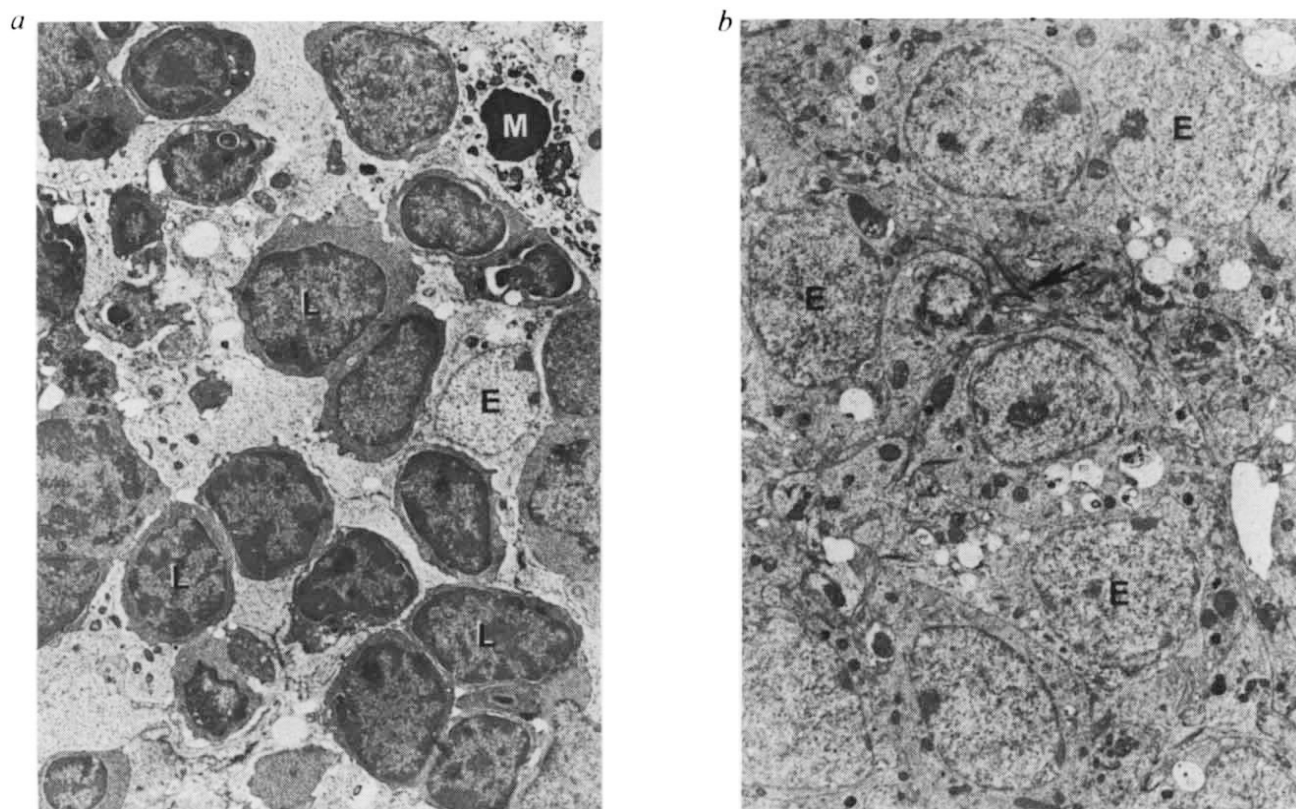


Fig. 1 *a*, Day-14 fetal CBA thymus cultured for 5 days without dGuo. Note the marked heterogeneity of cell types, with numerous lymphocytes (L) surrounding epithelial cells (E) and occasional macrophages (M). $\times 900$. *b*, CBA fetal thymus cultured for 5 days in media containing 1.35 mM dGuo. Note the homogeneous network of epithelial cells containing characteristic tonofilaments (arrow); desmosomes and prominent Golgi were visible under higher magnification. Note the absence of lymphocytes and macrophages. $\times 900$.

Methods. Day-14 fetal thymuses were cultured on Nucleopore disks placed on collagen sponges (Helitrex) resting in tissue culture medium. Medium consisted of RPMI 1640 (Gibco) with 10% fetal bovine serum, antibiotics (penicillin, streptomycin and Fungizone), glutamine, 5×10^{-5} M 2-mercaptoethanol, and 10 mM HEPES. dGuo (Sigma) was added to a final concentration of 1.35 mM. For electron microscopy, thymuses were fixed in modified Karnovsky's fixative and 1% osmium tetroxide, embedded in Epon, then sectioned and stained with uranyl acetate and lead citrate. Sections were viewed with a Hitachi HU 12A electron microscope.

grafted mice were constructed by transplanting B6 or CBA thymus grafts to adult thymectomized (ATx) F_1 mice, exposing the mice to irradiation (1,100 rad), usually 2 weeks after thymus grafting, and then reconstituting the mice with F_1 marrow cells. Mice were primed to KLH 2 months post-irradiation. As shown in group *a* of Table 1, purified T cells prepared from the draining lymph nodes (LN) of control $F_1 \rightarrow F_1$ chimaeras responded to KLH presented by cells of either parental strain, whereas T cells from $F_1 \rightarrow$ parent chimaeras showed marked restriction to thymic (host) H-2 determinants. Thus, $F_1 \rightarrow B6$ ($H-2^b$) chimaeras gave strong proliferative responses to antigen presented by $H-2^b$ (B6) cells but only minimal responses with $H-2^k$ (B10.BR) antigen-presenting cells (APC) (group *b*). Similarly, $F_1 \rightarrow CBA$ ($H-2^k$) chimaeras responded well to KLH presented by $H-2^k$ APC but not by $H-2^b$ APC (group *f*). Similar strong restriction to thymic H-2 determinants was observed with T cells taken from ATx F_1 mice grafted with thymus from B6 or CBA mice (groups *c, g*). To minimize the number of thymus-derived MØ/DC in the grafts, the thymuses were taken from 14-day-old fetuses. (To our knowledge, all previous studies dealing with the effects of thymus grafting on H-2 restriction used either neonatal or adult grafts.)

Although the early fetal thymus does contain small numbers of MØ/DC¹⁹, the enormous (>30-fold) enlargement that occurs when the fetal thymus is grafted is difficult to reconcile with the view that MØ/DC of thymus origin imprint H-2 restriction. To sustain this view, one has to argue that the few MØ/DC present

in the thymus at the time of grafting (1) underwent substantial clonal expansion after grafting and (2) prevented entry of host-derived cells. Concerning the latter point, in contrast to the findings of Longo *et al.*^{4,5}, we have observed that host cells enter thymus grafts in considerable numbers within 2 weeks (the time at which the grafted mice were irradiated). The data in Table 2 demonstrate that this influx of host-derived cells includes cells with APC function, that is, cells able to present antigen to a KLH-specific T-cell clone. The clone, C.I.e., is restricted by I-E^k molecules and responds to antigen presented by (B6 $\times$ CBA) F_1 but not B6 APC. It can be seen that, unlike B6 thymocytes, strong APC function for the clone was found with thymocytes from (1) 16-day-old fetal B6 thymuses 'parked' for 2 weeks in normal F_1 mice and (2) thymuses taken from $F_1 \rightarrow B6$ chimaeras 2 weeks after irradiation and marrow reconstitution. No stimulation was found with thymuses from reciprocal B6 $\rightarrow F_1$ chimaeras, which implies that thymic APC are almost entirely replaced by immigrant donor cells within 2 weeks of irradiation. Other groups have reported similar findings²⁰⁻²². To counteract the objection that host-type cells take longer than 2 weeks to reach the particular site in the thymus that imprints restriction, some thymus-grafted mice were left for 2 months before irradiation. Again, T cells from these mice showed marked restriction to thymic H-2 determinants (Table 1, group *d*).

To obtain direct evidence on whether MØ/DC are necessary for inducing H-2 restriction, 14-day-old thymuses were cultured with dGuo for 5 days *in vitro*. In accord with the studies of

Table 2 Influx of bone marrow-derived APC into thymuses

APC source	³ H-TdR incorporation (c.p.m. × 10 ³) by clone C.l.e	
	–KLH	+5 µg ml ^{–1} KLH
Normal (B6 × CBA)F ₁ thymus	0.4 (±0.3)	14.7 (±1.1)
Normal B6 thymus	0.3 (±0.1)	0.5 (±0.2)
16-day B6 fetal thymus grafted 2 weeks in normal F ₁	0.2 (±0.2)	11.2 (±0.9)
F ₁ → B6 thymus 2 weeks after irradiation (1,100 rad) and reconstitution	0.5 (±0.1)	8.3 (±0.8)
B6 → F ₁ thymus 2 weeks after irradiation (1,100 rad) and reconstitution	0.2 (±0.01)	0.4 (±0.1)

To test the origin of APC in the thymus of chimaeras, thymus cell suspensions were fractionated on a discontinuous Percoll gradient²⁹. Most cells with APC function were found in the specific gravity 1.05–1.06 and 1.06–1.07 interfaces. Cells pooled from these fractions (2–4% of total thymocytes) were irradiated (1,500 rad) and placed in culture (5 × 10⁵ cells per well) with a T-cell clone, C.l.e, specific for KLH seen in the context of 1-E^k molecules, the C.l.e. clone, added to the cultures at 10⁵ cells per well, was derived from an F₁ → B6 chimaera primed with KLH. Soluble KLH was added to cultures where indicated. Cultures were pulsed with 1 µCi ³H-TdR on day 2 and collected on day 3. Data are mean (±s.d.) of triplicate cultures. B6 → F₁ thymus APC, while unable to present KLH to clone C.l.e., are able to stimulate CBA lymph node T cells in mixed-lymphocyte cultures, though weakly (data not shown).

Jenkinson *et al.*⁸, electron microscopy showed that dGuo-treated thymuses contained almost no lymphocytes or macrophages and consisted almost entirely of epithelial cells (Fig. 1) (confirmed in all of three experiments with two B6 and one CBA dGuo-treated thymuses). Despite the extreme paucity of MØ/DC in these thymuses, T cells differentiating in dGuo-treated thymus grafts exhibited strong restriction to thymic H-2 determinants (Table 1, groups *e*, *h*).

Although we cannot at present explain the contradictory data of Longo *et al.*^{4,5}, the above results seem to constitute a strong case for epithelial cells rather than MØ/DC are responsible for imprinting the H-2-restricted specificity of T_H cells; using a very similar approach to that of Longo and Schwartz⁴, Zinkernagel²³ and Fink and Bevan²⁴ found no evidence that MØ/DC induce H-2 restriction of T_C cells. In the case of tolerance induction, however, there is accumulating evidence from both *in vivo*^{6,7} and *in vitro*^{25–27} studies that only MØ/DC and not epithelial cells induce tolerance across H-2 barriers. The only evidence against this notion is that tolerance to thymic H-2 determinants has been observed with T cells differentiating in thymuses precultured at low temperature²⁸, another procedure that selectively destroys bone marrow-derived cells. It remains possible, however, that culture at low temperature is less effective than dGuo treatment in removing MØ/DC. Note that, in agreement with the findings of others⁶, we have confirmed that allografts of dGuo-treated fetal thymuses are accepted for prolonged periods and do not induce tolerance as measured by mixed-lymphocyte reactions (unpublished data), implying that our dGuo-treated thymus grafts were indeed depleted of MØ/DC.

According to the view that different cell types in the thymus control the induction of tolerance as opposed to H-2 restriction, T-cell specificity might be shaped as follows. Receptor-bearing thymocytes make initial contact with H-2 determinants on epithelial cells in the cortex, a region largely free of MØ/DC. After 'learning' self-H-2-restricted specificity, the T cells then migrate from the cortex to the cortico-medullary junction. Here, contact with MØ/DC deletes those T cells expressing high affinity for self-H-2 determinants. Low-affinity T cells then pass through this filter and migrate into the periphery.

We thank Dr C.-M. Chang for the electron microscopy studies. This work was supported in part by USPHS grants AI21687,

CA38355, CA35048 and CA25803. D.L. is supported by a NIH Medical Scientist Training Program grant (5-T32-GM-07170) to the University of Pennsylvania.

Received 22 October; accepted 9 December 1985.

1. Klein, J. (ed.) *Immunology: The Science of Self-Nonself Discrimination* 1–687 (Wiley, New York, 1982).
2. Zinkernagel, R. M. *Immun. Rev.* **42**, 224–270 (1978).
3. Bevan, M. J. & Fink, P. J. *Immun. Rev.* **42**, 3–19 (1978).
4. Longo, D. L. & Schwartz, R. H. *Nature* **287**, 44–46 (1980).
5. Longo, D. L. & Davis, M. L. *J. Immun.* **130**, 2525–2527 (1983).
6. Ready, A. R., Jenkinson, E. J., Kingston, R. & Owen, J. J. T. *Nature* **310**, 231–233 (1984).
7. von Boehmer, H. & Schubiger, K. *Eur. J. Immun.* **14**, 1048–1052 (1983).
8. Jenkinson, E. J., Franchi, L., Kingston, R. & Owen, J. J. T. *Eur. J. Immun.* **12**, 583–587 (1982).
9. Sprent, J. *J. exp. Med.* **147**, 1838–1842 (1978).
10. Kappler, J. W. & Marrack, P. *J. exp. Med.* **148**, 1510–1522 (1978).
11. Singer, A., Hathcock, K. S. & Hodes, R. J. *J. exp. Med.* **153**, 1286–1301 (1981).
12. von Boehmer, H. & Haas, W. *J. exp. Med.* **150**, 1134–1142 (1979).
13. Bevan, M. J. *Nature* **269**, 417–418 (1977).
14. Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 882–896 (1978).
15. Matzinger, P. & Mirkwood, G. *J. exp. Med.* **148**, 84–92 (1978).
16. Kruisbeek, A. M., Sharrow, S. O. & Singer, A. *J. Immun.* **130**, 1027–1032 (1983).
17. Wagner, H. *et al. J. exp. Med.* **153**, 1517–1532 (1981).
18. Corradin, G., Etlinger, H. M. & Chiller, J. M. *J. Immun.* **119**, 1048–1053 (1977).
19. Robinson, J. H. *J. Immun.* **130**, 1592–1595 (1983).
20. Barclay, A. N. & Mayrhofer, G. *J. exp. Med.* **153**, 1666–1671 (1981).
21. Kyewski, B. A., Fatham, C. G. & Kaplan, H. S. *Nature* **308**, 196–199 (1984).
22. Jordan, R. K., Bentley, A. L., Perry, G. A. & Crouse, D. A. *J. Immun.* **134**, 2155–2160 (1985).
23. Zinkernagel, R. M. *J. exp. Med.* **156**, 1842–1847 (1982).
24. Fink, P. J. & Bevan, M. J. in *Recognition and Regulation in Cell Mediated Immunity* (eds Watson, J. D. & Marbrook, J.) 107–125 (Dekker, New York, 1985).
25. Robinson, J. H. & Owen, J. J. T. *Nature* **271**, 758–760 (1978).
26. Good, M. F., Pyke, K. W. & Nossal, G. J. V. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3045–3049 (1983).
27. Jenkinson, E. J., Jhittay, P., Kingston, R. & Owen, J. J. T. *Transplantation* **39**, 331–333 (1985).
28. Jordan, R. K., Robinson, J. H., Hopkinson, N. A., House, K. C. & Bentley, A. L. *Nature* **314**, 454–456 (1985).
29. Nathanson, S. D., Zamfirescu, P. L., Drew, S. I. & Wilbur, S. J. *immun. Meth.* **18**, 225–234 (1977).

Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy

Klas Kärre*, Hans Gustaf Ljunggren*, Gerald Piontek* & Rolf Kiessling†

Departments of *Tumor Biology and †Immunology, Karolinska Institute, Stockholm, Sweden

Metazoan organisms may discriminate between self and non-self not only by the presence of foreign antigens but also by the absence of normal self markers¹. Mammalian adaptive immune responses use the first strategy, with the additional requirement that foreign antigens are recognized in the context of self-major histocompatibility complex (MHC) products at the cell surface². Aberrant cells which fail to express MHC products adequately can therefore avoid detection^{2–4}. A more primitive but complementary defence system, eliminating such cells on the basis of absent self-markers, is suggested by a re-interpretation^{5,6} of phenomena associated with metastasis and natural resistance. We now show that murine lymphoma cells selected for loss of H-2 expression are less malignant after low-dose inoculation in syngeneic hosts than are wild-type cells, and that the rejection of such cells is non-adaptive. On the basis of our data, we suggest that natural killer cells are effector cells in a defence system geared to detect the deleted or reduced expression of self-MHC.

More specifically, the hypothesis^{5,6} stated that self-MHC genes influence the expression or recognition of inhibitor signals in interactions between potential target cells and natural effector cells. Our experimental approach was to test the prediction that selection for loss of H-2 in tumour lines should be accompanied by an increased sensitivity to natural resistance *in vitro* and *in vivo*. Three independent anti-H-2-resistant sublines from two lymphomas were selected. Such variant cells not only were able to escape adaptive alloreactivity *in vitro* and *in vivo*, but, as predicted by our hypothesis, had also become unable to grow

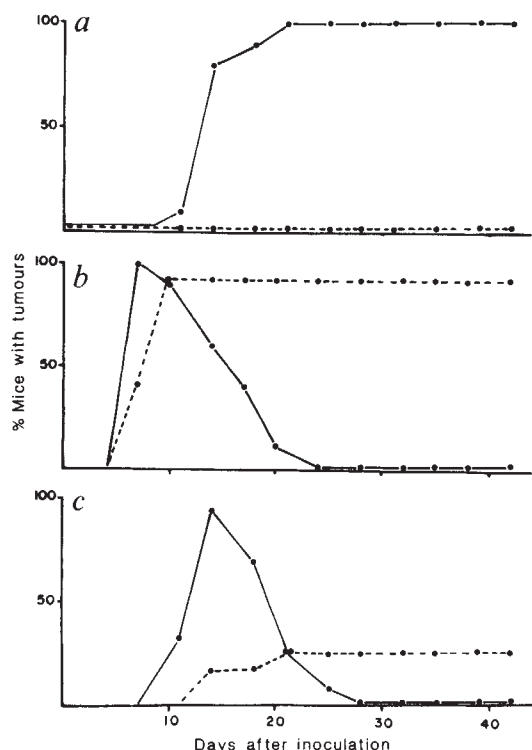


Fig. 1 Proportion of syngeneic and allogeneic mice with palpable tumours at different time points after subcutaneous injection of RMA (RBL-5, mutagenized but not selected; solid line) and RMA H-2 sel. (mutagenized, selected, H-2-resistant subline; dashed line). *a*, 10^4 cells in syngeneic C57BL mice; summary of two experiments with a total of 12 mice for each tumour. *b*, 10^6 cells in allogeneic A.BY mice; summary of four experiments with a total of 18–21 mice for each tumour. *c*, 10^3 cells in allogeneic A.BY mice; summary of four experiments with a total of 18–21 mice for each tumour.

out from small inocula in the normal syngeneic host, and switched from the natural killer (NK)-resistant to the NK-sensitive phenotype as a consequence of the selection against H-2 expression.

We used the RBL-5 lymphoma (H-2^b) for these studies because it readily expresses H-2 and is highly malignant in the normal syngeneic host. In contrast to two other C57BL-derived lymphomas, the number of tumour takes after a threshold inoculum was not significantly increased in NK-cell-defective mice⁷. This lymphoma is therefore probably regarded as close to 'normal self' by natural resistance mechanisms in the syngeneic or H-2^{b/b}-compatible hosts. In contrast, the subline that had been selected for loss of H-2 expression (RMA H-2 sel) failed to grow in such mice, even after a 10-fold increase in the dose required for 100% takes of the wild-type tumour (Table 1). The subline had been derived after mutagenization and repeated treatment with alloantiserum and complement. It was completely resistant to this treatment, due to a selective loss or reduction in expression of MHC class I gene products (Table 1) and β_2 -microglobulin; it did not stain significantly above normal control serum with alloantibodies on analysis by fluorescence-activated cell sorter (FACS) (Fig. 1), while binding and complement-dependent killing with anti-Thy 1.2 antibodies were unaffected (Table 1). The wild-type and the selected line were both negative for monoclonal antibodies against TL and I-A^b antigens (not shown).

In view of the proposed hypothesis, it was important first to study natural resistance in syngeneic (C57BL) or at least semi-syngeneic (A.BY $\times$ C57BL) H-2^{b/b} homozygous hosts, in order to avoid combinations producing strong T-cell reactivity. In an attempt to obtain the latter situation in subsequent studies,

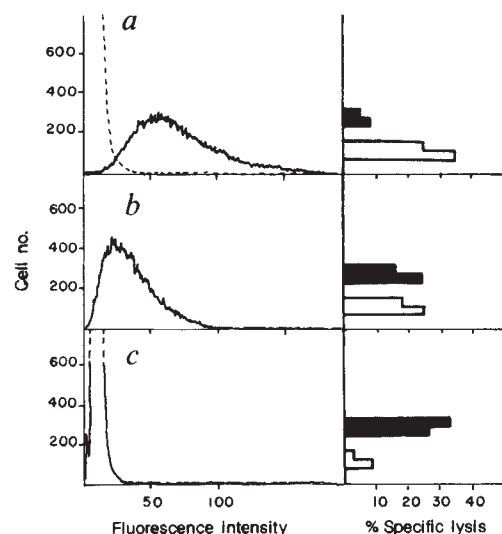


Fig. 2 H-2 expression, susceptibility to allospecific cytotoxic T lymphocytes (CTL), and poly-I:C-induced NK cells of RBL-5 sublines established after mutagenization, before selection (= RMA, *a*), after two selections with alloantiserum + complement (*b*) and after the completed series of five selections (= RMA H-2 sel., *c*). Left-hand panels show H-2^b expression by FACS analysis after indirect staining with hyperimmune A/Sn anti-A.BY alloantiserum (1/20) and fluorescent conjugated rabbit anti-mouse immunoglobulin (Dakopatts). The fluorescence profiles were determined on a FACS IV using a laser light output of 200 mV, photomultiplier at 450 V and fluorescence gain $\times 2$. Nonspecific background staining, estimated by using the reciprocal antiserum A.BY anti-A/Sn (anti-H-2^a) as the first antibody, was similar for the three lines and is shown here for the nonselected line only (dashed curve, *a*). Right-hand panels show the sensitivity of the corresponding cell lines to anti-H-2^b cytotoxic T cells derived from A/Sn anti-A.BY mixed leukocyte cultures, and tested at an effector/target ratio of 6/1 (open bars, with the upper and lower part corresponding to the same two independent tests for all cell lines) and splenocytes from H-2-syngeneic C57BL or (A.BY $\times$ C57BL)_{F1} mice injected with poly-I:C (150 μ g intraperitoneally 18 h before test), tested at an effector/target ratio of 100/1 (solid bars, with the upper and lower part corresponding to the same two independent tests for each of the cell lines). The mixed lymphocyte cultures and the 4–6-h ⁵¹Cr-release assays were set up as described previously³⁹ and all three lines were tested in parallel in each of the experiments.

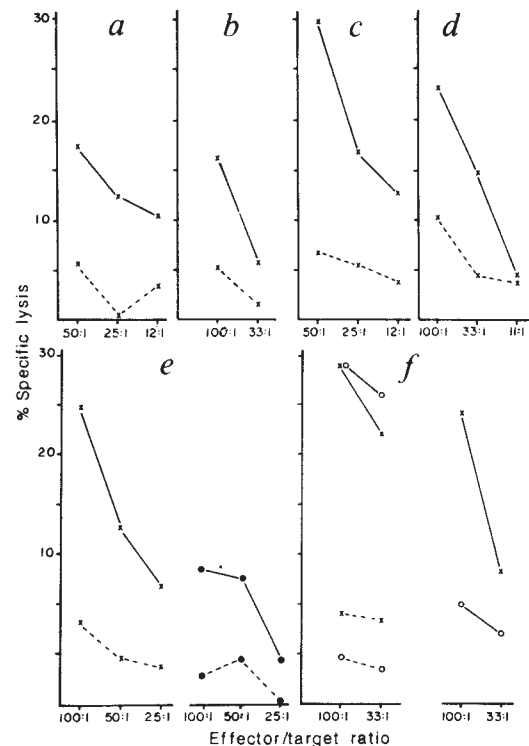
high-dose inocula were given to inbred A.BY mice (rather than outcrossed to C56BL), presenting a minor histocompatibility barrier to the RBL-5 lymphoma⁸. The pattern of progressive tumour growth then changed to the opposite: H-2-variant lymphoma cells grew progressively once they had formed a tumour, whereas all wild-type tumours were rejected after temporary growth (Fig. 1). High-dose challenge (1×10^6 cells) of preimmunized syngeneic C57BL mice gave a similar picture: complete protection was seen with the wild-type line, while H-2-resistant cells failed to immunize against themselves (Table 1). These results and observations on tumour outgrowth in pre-irradiated hosts (Table 1) indicated that the reduced tumorigenicity of H-2-lacking variant cells was not due to an intrinsic inability to grow *in vivo*, but rather to an active host defence, dependent on radiosensitive (precursor) cells, but independent of tumour immunogenicity in the classical sense. It was important to establish that this was the case, in view of reports that low malignant mutant or transfectant tumour lines can acquire the ability to elicit strong adaptive T-cell responses^{9–12}. Indeed, the variants seemed to escape T-cell recognition *in vivo*, probably because minor histocompatibility antigens and tumour antigens are recognized only in the context of self-MHC products^{2,13} and thus fulfilled one of the criteria for action by the postulated defence system. The variant cells were also able to escape anti-H-2^b

Table 1 Summary of experimental results

Sublines			Serology						Tumour growth <i>in vivo</i>							
Original line	Mutagenization series selection	Designation	A/Sn → A.BY allosensitiser		20-8-4S K ^D mab		β_2 M mab	β_2 M rabbit	F7DG Thy 1.2 mab		Untreated				Pre-immunized 10 ⁶	Pre-irradiated 10 ⁴
			CX	FACS	CX	FACS			CX	FACS	10 ³	10 ⁴	10 ⁵	10 ⁶		
RBL-5	A—none	RMA	++	++	++	++	+	++	++	++	7/7	16/16	10/10	9/9	1/9	9/9
RBL-5	A—A/Sn → A.BY	RMA H-2 sel.	—	—	—	(±)	—	(±)	++	++	0/9	0/14	NT	5/5	5/5	4/4
RBL-5	B—A/Sn → A.BY	RMB H-2 sel.	—	—	—	(±)	—	(±)	++	++	1/14	0/16	1/12	4/5	5/5	5/5
EL-4	A—none	EL-4MA	++	++	++	++	NT	NT	NT	NT	3/10	6/10	7/10	NT	NT	NT
EL-4	A—A/Sn → A.BY	EL-4MA H-2 sel.	—	(±)	—	(±)	NT	NT	NT	NT	0/10	1/13	1/14	NT	NT	NT

Sublines: RBL-5 is a Rauscher virus-induced lymphoma and EL-4 a benzpyrene-induced leukaemia, both originating in the C57BL strain. The cells, maintained in RPMI with 10% fetal calf serum and penicillin-streptomycin, were mutagenized with ethanemethanesulphonate (EMS, Sigma), 200 $\mu\text{g ml}^{-1}$ for 24 h. After washes, the cells were allowed to recover for 1 week and then treated with allosensitiser (A/Sn → A.BY diluted 1:20, 10^7 cells ml^{-1}) followed by rabbit complement (Pel Freez, 1/10) for 1–4 h. The cells were cultured without any attempts to remove the dead cells, and the survivors (<0.25%) were expanded. The treatment was repeated at least four times. **Serology:** The A/Sn → A.BY (anti-H-2^b) serum was produced in our animal colony at the Karolinska Institute. The monoclonal antibody (mab) 20-8-4S (ref. 38) was obtained from the American Type Culture Collection. The anti- β_2 -microglobulin (β_2 M) reagents were kindly sent by Drs Shen and Boyse (Sloan Kettering Memorial Institute, New York) (mab) and Dr Tanigaki (Roswell Park Memorial Institute, Buffalo) (rabbit antiserum). The F7D5 reagent was a gift from Dr E. Clark (Genetic Systems, Seattle). CX, complement-dependent cytotoxicity in a two-step assay evaluated by trypan blue exclusion. ++, 95–100% of cells killed, reciprocal titre no less than 100. +, 70–95% of cells killed, reciprocal titre 10–50; —, below background killing with complement alone. FACS, indirect fluorescence, after a two-step staining procedure using a FACS IV (Becton & Dickinson) with a laser output of 200 mV, photomultiplier at 600 V and a constant fluorescence gain for each antibody. The settings used gave a background median fluorescent intensity (MFI) below 10 on a linear scale (using normal mouse of rabbit serum, C57BL antisera to other H-2 haplotypes or T-lymphoma supernatant instead of the specific antibody in the first incubation). ++, MFI > 100 under the same conditions; +, MFI 20–100; (±) MFI, 5–19; —, below background. Secondary reagents: fluorescein isothiocyanate-conjugated rabbit-anti-mouse immunoglobulin or swine anti-rabbit immunoglobulin (Dakopatts). The cells were incubated at 10^7 ml^{-1} , 30 min on ice in both steps. **Tumour growth *in vivo*:** All tests were with ascites tumours, serially passaged in 400R-pre-irradiated mice to avoid immunoselection *in vivo*. Different numbers of tumour cells were injected subcutaneously to groups of three to six H-2-syngeneic hosts, either C57BL or (A.BY × C57BL)_{F1} for all tumours in each experiment. The two types of mice gave consistent results, and the number of mice with tumour takes (out of totally injected) have therefore been pooled from several different experiments, each one including wild-type and variant cells injected to parallel groups. Two experiments with 10^4 cells included parallel groups of preirradiated mice (400R, 24 h before tumour challenge). The data with 10^6 injected cells represent two experiments where half of the mice had been preimmunized with the challenging tumour (three weekly injections of 2×10^7 10,000 R irradiated RMA or RMA H-2 sel. cells). NT, not tested.

Fig. 3 Susceptibility of wild-type and H-2-selected sublines to killing by spleen cells from poly-I:C-treated (150 μg per mouse intraperitoneally 18 h before injection) or untreated mice. *a, b*, Two independent experiments with effector cells from poly-I:C-treated C57BL mice tested against EL-4MA (dashed line) and EL-4MA H-2 sel (solid line). *c, d*, Two independent experiments with effector cells from poly-I:C-treated C57BL (*c*) and (A.BY × C57BL)_{F1} (*d*) mice tested against RBL-5 wild-type (dashed line) and RMB H-2 sel. (solid line). *e*, Nylon wool column-passed³⁵ spleen cells from normal C57BL (×) and A.BY (○) mice tested against RMA (dashed line) and RMA H-2 sel. (solid line) in a 16-h ⁵¹Cr release assay. *f*, Left, effector cells from poly-I:C-treated (A.BY × C57BL)_{F1} mice tested against RMA (dashed lines) and RMA H-2 sel. (solid lines), with (○) and without (×) pretreatment of targets with β -interferon as described above. Median H-2 fluorescence intensity by FACS analysis: RMA, 105, β -interferon-treated RMA, 180; RMA H-2 sel. < 15 before as well as after β -interferon treatment; YAC-1, 28; β -interferon-treated YAC-1, 71 (staining procedure as in Fig. 2, except that A.BY anti-A/Sn H-2^a serum was used as the first antibody with YAC-1 cells; background was 6–11 for all lines. All cytotoxicity tests were performed with the ⁵¹Cr-release assay, as described previously^{39,40}, with an incubation time of 4–6 h, except in *e* (16 h).



cytotoxic lymphocytes *in vitro* (Fig. 2) as a result of the selection. However, at the same time, they had switched from the NK-resistant to the NK-sensitive phenotype (Fig. 2).

Although the *in vivo* results presented here do not prove definitively that the natural resistance was mediated by NK cells, the non-adaptive nature of the rejection and the results on irradiated hosts are consistent with this interpretation. The dose of 400R reduces splenic cellularity by 95%; even if some functional NK cells remain, the total activity in terms of lytic units per animal is reduced within 24 h, and falls to zero within the following weeks, that is, the period between tumour injection and outgrowth in our experiments¹⁴. This enhanced sensitivity to natural resistance *in vitro* and *in vivo* was observed with an independently derived H-2-resistant RBL-5 subline, as well as

with an EL-4 subline selected in a similar fashion (Table 1, Fig. 3). The general lysability of the wild-type control line was also confirmed in the case of EL-4, which was efficiently killed (>50% lysis at effector target/ratio 5:1) by anti-H-2-specific, mixed lymphocyte culture-derived effector cells (data not shown).

According to the proposed model, the tumour cells would be unable to inhibit triggering of NK lysis after complete or partial loss of self H-2^{5,6}. This would be consistent with several observations on NK cells, including high sensitivity of *in vitro*-grown YAC-1 lymphoma cells, immature thymocytes and teratocarcinoma cells, all of which have low or defective MHC expression^{15,16}. The corresponding cells show reduced sensitivity when H-2 is induced (by differentiation^{17,30}, interferon¹⁹ or

growth *in vivo*²). NK-cell-mediated rejection of H-2-disparate bone marrow and lymphoid grafts, including the F₁ hybrid resistance phenomenon^{18,21,22} and early elimination of allogeneic lymphocytes²³, are other examples that could be reinterpreted to reflect recognition of incomplete self-H-2 expression: allogeneic grafts lack H-2 genes of the recipient, and parental grafts fail to match one haplotype, that is, they lack 50% or more of the total 'self'-H-2 molecules expressed in the F₁ recipient environment. This predicts that the relationship between reduced self expression and NK susceptibility should be quantitative rather than dependent on complete deletion of H-2 products, as is indeed supported by findings with a subline of RBL-5 derived after the two first selections with anti-H-2 + complement. These cells had an average loss on H-2 fluorescence corresponding to more than 50% of the total H-2 expression in the original line, and a clear increase in NK sensitivity (Fig. 2).

The hypothesis would also explain why it is difficult to select H-2⁻ variants *in vivo*²⁴, and predict a selection for increased H-2 expression during tumour progression. The correlation between metastatic potential and increased expression of all²⁵ or certain (D-end encoded^{26,27}) MHC class I alleles observed for several experimental tumours could therefore be reinterpreted as selection of phenotypes able to escape NK cells. However, preferential outgrowth of cells with reduced or deleted expression of certain MHC alleles (often H-2K-end encoded) has been observed after large-dose inoculation ($\geq 10^5$ cells) of some virally induced, strongly antigenic tumours^{12,28,29}. T cells presumably represented the major surveillance mechanism in those systems, as they would in the present experiments, where H-2-deficient sublines grew out preferentially after transplantation of large-dose inocula to preimmunized C57BL or histoincompatible A.BY mice (Table 1, Fig. 1).

Burnet pointed out the fundamental difference between strategies for self-non-self-discrimination in invertebrates and mammals¹. In the colonial tunicate *Botryllus*, rejection of allogeneic cells as well as control against self-fertilization is thought to operate by a mechanism scanning for the presence or absence of self-markers, encoded by a single locus with considerable polymorphism^{1,31}. Such a system may also have been fixed in mammals, despite the development of adaptive immunity. The selective pressures would have required not only a back-up system eliminating aberrant (MHC loss) cells escaping detection by T lymphocytes, but also a rapid first-line defence with a certain selectivity. Examples exist of lytic as well as transforming viruses that cause a profound but selective reduction of H-2 expression in infected cells^{29,32}. The latter would be eliminated by the proposed defence mechanism, while sparing of normal cells in the tissues would be guaranteed by an increase in their H-2 expression due to interferon stimulation early in immune responses and NK-cell activation. Interferons do reduce the NK susceptibility of many targets³³; the mechanisms involved are unknown, but may be related to MHC gene regulation as suggested by observations on variant cells in this (Fig. 3f) and another study⁴¹.

The present findings may also be relevant to previous speculations by Snell³⁴ and Dausset³⁵ concerning the primordial functions of MHC genes, that is, functions that would not be related to T-cell restriction. The possibility of a common regulation of H-2 expression, NK-cell-resistant phenotype and malignancy properties in lymphoma cells is also very interesting in relation to recent advances in the research on regulation of MHC class I genes, particularly in tumour cells^{36,37}.

This study was supported by PHS grants 5 ROI CA 25250-06 and 5 ROI CA 26782-06 and by the Swedish Cancer Society. We thank Ms Maj-Lis Solberg, Margareta Hagelin, Erne Eriksson, Gunnel Brolin and Eva Lotta Jonsson for technical assistance, Mr Anders Carstenson for advice and assistance with the FACS analysis, Ms Inger Lindfors for secretarial assistance, and Professors George Klein and Hans Wigzell for stimulating discussions and support.

Received 7 August; accepted 9 December 1985.

1. Burnet, F. M. *Nature* **232**, 230-235 (1971).
2. Zinkernagel, R. M. & Doherty, P. C. *J. exp. Med.* **141**, 1427-1436 (1975).
3. Doherty, P. C., Knowles, B. B. & Wettstein, P. J. *Adv. Cancer Res.* **42**, 1-66 (1984).
4. Sanderson, A. R. & Beverley, P. C. L. *Immun. Today* **4**, 211-213 (1983).
5. Kärre, K. thesis, Karolinska Inst. (1981).
6. Kärre, K. in *Mechanisms of Cytotoxicity by Natural Killer Cells* (eds Herberman, R. B. & Callewaert, D.) 81-91 (Academic, Orlando, 1985).
7. Kärre, K., Klein, G. O., Kiessling, R., Klein, G. & Roder, J. C. *Int. J. Cancer* **26**, 789-797 (1980).
8. Klein, J. in *Biology of the Mouse Histocompatibility-2 Complex*, 36-37 (Springer, New York, 1975).
9. Boon, R. *Adv. Cancer Res.* **39**, 121-151 (1983).
10. Frost, P., Kerbel, R. S., Bauer, E., Tartanella-Biondo, R. & Cefalu, W. *Cancer Res.* **43**, 125-132 (1983).
11. Hui, K., Grosvels, F. & Festenstein, H. *Nature* **311**, 750-752 (1984).
12. Hämmerling, G. J. *et al. Nature* **315**, 301-306 (1985).
13. Bevan, M. J. *J. exp. Med.* **142**, 1349-1364 (1975).
14. Cudkovic, G. & Hochman, P. S. *Immun. Rev.* **44**, 13-41 (1979).
15. Cikes, M., Friberg, S. Jr & Klein, G. *J. natn. Cancer Inst.* **30**, 347-361 (1973).
16. Stern, P. *et al. Nature* **285**, 341-342 (1982).
17. Hansson, M. *et al. J. Immun.* **123**, 765-771 (1979).
18. Carlson, G. A., Melnychuk, D. & Meeker, M. J. *Int. J. Cancer* **25**, 111-122 (1980).
19. Hansson, M., Kiessling, R., Andersson, B. & Welsh, R. M. *J. Immun.* **125**, 2225-2231 (1980).
20. Becker, S., Kiessling, R., Lee, N. & Klein, G. *J. natn. Cancer Inst.* **61**, 1493-1498 (1978).
21. Cudkovic, G. & Bennet, M. *J. exp. Med.* **134**, 1513-1528 (1971).
22. Klein, G. O., Klein, G., Kiessling, R. & Kärre, K. *Immunogenetics* **6**, 561-569 (1978).
23. Rolstad, B. & Ford, W. L. *Immun. Rev.* **73**, 87-114 (1983).
24. Möller, E. *J. natn. Cancer Inst.* **33**, 979-987 (1964).
25. Haywood, G. R. & McKhann, C. *J. exp. Med.* **133**, 1171-1187 (1971).
26. Katrav, S., De Baetselier, P., Tarbakovsky, B., Feldman, M. & Segal, S. *J. natn. Cancer Inst.* **71**, 317-324 (1983).
27. Eisenbach, L., Segal, S. & Feldman, M. *Int. J. Cancer* **32**, 113-120 (1983).
28. Gooding, L. R. *J. Immun.* **129**, 1306-1312 (1982).
29. Schrier, P. I., Bernards, R., Vaessen, J., Houweling, A. & van der Eb, A. J. *Nature* **305**, 771-775 (1983).
30. Gidlund, M. *et al. Nature* **292**, 848-850 (1981).
31. Scofield, V. L., Schlumpberger, J. M., West, L. A. & Weissman, I. L. *Nature* **295**, 499-502 (1982).
32. Hecht, T. & Summers, D. J. *Virology* **10**, 378-385 (1972).
33. Welsh, R. M., Kärre, K., Hansson, M., Kunkel, L. A. & Kiessling, R. *J. Immun.* **126**, 219-225 (1981).
34. Snell, G. *Transplant Proc.* **8**, 147-156 (1976).
35. Dausset, J. *Dev. comp. Immun.* **5**, 1-4 (1981).
36. Brickell, P. M., Latchman, D., Murphy, D., Wilson, K. & Rigby, P. W. J. *Nature* **306**, 756-761 (1984).
37. Baldacci, P., Pozo, F., Gisselbrecht, S. & Kourilsky, P. J. *exp. Med.* **158**, 1294-1306 (1981).
38. Ozato, K. & Sachs, D. H. *J. Immun.* **126**, 317-321 (1981).
39. Kärre, K., Seeley, J., Eriksson, E., Burton, R. & Kiessling, R. *J. exp. Med.* **156**, 385-403 (1983).
40. Julius, M. H., Simpson, E. & Herzenberg, L. A. *Eur. J. Immun.* **3**, 645-651 (1973).
41. Piontek, G. *et al. J. Immun.* **135**, 4281-4288 (1985).

Expression of apamin receptor in muscles of patients with myotonic muscular dystrophy

Jean-François Renaud, Claude Desnuelle*, Heidy Schmid-Antomarchi, Michel Hugues, Georges Serratrice* & Michel Lazdunski

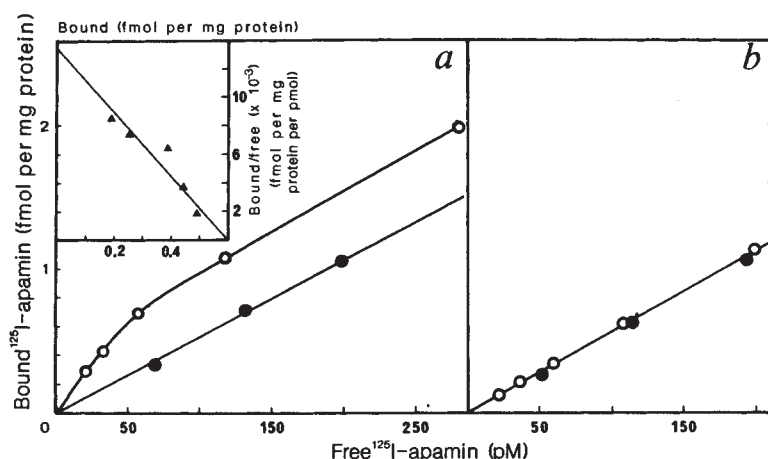
Centre de Biochimie du Centre National de la Recherche Scientifique, Parc Valrose, 06034 Nice Cedex, France

*Clinique Rhumatologique et des Maladies Neuromusculaires, Centre Hospitalo-Universitaire de la Timone, 13385 Marseille Cedex 4, France

Myotonic muscular dystrophy, or Steinert disease, is a dominantly inherited disease of muscle which occurs with a frequency of between 1 in 18,000 and 1 in 7,500 people (refs 1, 2). One of the prominent clinical manifestations is muscle stiffness and difficulty in relaxation of muscles after voluntary contractions. Electrophysiological signs of myotonia include increased excitability with a tendency to fire trains of repetitive action potentials in response to direct electrical and mechanical stimulation. Most experimental and clinical data suggest that myotonic muscular dystrophy arises from genetically induced alterations of the muscle membrane³. We show here for the first time that muscle membranes of patients with myotonic muscular dystrophy contain the receptor for apamin, a bee venom toxin known to be a specific and high-affinity blocker of one class of Ca²⁺-activated K⁺ channels in mammalian muscle⁴⁻⁶. The apamin receptor is completely absent in normal human muscle as well as in muscles of patients with spinal anterior horn disorders.

Fig. 1 Binding of ^{125}I -apamin in: *a*, a deltoid muscle of a 26-yr-old patient with myotonic muscular dystrophy; *b*, a normal human muscle. Total ($\circ$) and nonspecific ($\bullet$) binding were determined in parallel in the absence and in the presence of an excess of unlabelled apamin (1 μM), respectively. A Scatchard plot of the specific ^{125}I -apamin binding, calculated as the difference between total and nonspecific binding, is given in the inset of *a*.

Methods. Frozen biopsied muscles were thawed and washed with an ice-cold 20 mM Tris-HCl buffer, pH 7.4, containing 0.25 M sucrose and 1 mM EDTA (TSE buffer). Samples were dissected free from fat and connective tissues, minced, rinsed with ice-cold TSE buffer and homogenized at setting 5 in a Polytron homogenizer PT 10 using three 5-s bursts. Homogenates were centrifuged at 1,500g for 10 min at 4 °C, the pellets discarded and supernatants centrifuged at 100,000g for 15 min. The supernatants of the 100,000g fraction were centrifuged in fresh TSE buffer at a final concentration of 10 mg ml⁻¹ membrane protein and were used directly for binding assays. Protein concentrations of purified membranes were determined according to Hartree¹⁷. Binding assays of ^{125}I -apamin (2,000 Ci mmol⁻¹) to purified skeletal muscle membranes were carried out using the procedure described previously⁴. The equivalent of 0.4 mg protein was equilibrated in 1 ml standard incubation medium containing 100 mM Tris-HCl, 5.4 mM KCl and 0.1% bovine serum albumin (BSA) at pH 8.4 and 4 °C in the presence of increasing concentrations of ^{125}I -apamin. The incubation lasted for 60 min at 4 °C and was stopped by filtering in duplicate 400- μl aliquots of the incubation medium through Sartorius filters (SM 11107; 0.2- μm pore size). Filters were rapidly washed twice with 5 ml medium containing 20 mM Tris-HCl (pH 8.4) and 0.1% BSA.



Samples (0.1–0.5 g) of human muscle were obtained by open biopsy after informed consent of the patient, immediately frozen in liquid nitrogen and kept at –80 °C until used. One part of each sample was used for histochemical examination and the other for binding studies. The investigation described here concerns 11 patients between the ages of 17 and 69 years with symptomatic myotonic muscular dystrophy. Diagnosis was established by history, clinical examination and electromyography (EMG) as well as morphological and histochemical examination of muscle. Control samples of muscle were obtained from men and women undergoing surgery and showing no specific clinical features. Morphological, histochemical and EMG properties of these control muscles were normal.

The association of ^{125}I -apamin to membranes of myotonic muscular dystrophy muscles is shown by the equilibrium-binding data in Fig. 1*a*. The linearity of the Scatchard plot in Fig. 1*a* inset indicates that ^{125}I -apamin binds to a single class of receptor sites. The maximum binding capacity (B_{max}) and the dissociation constant (K_D) of the apamin–receptor complex are 0.6 fmol per mg protein and 51 pM, respectively. Table 1 summarizes the data obtained from muscles of 11 patients with symptomatic myotonic muscular dystrophy. Apamin receptors have been identified in muscle membranes for 8 of these patients. We obtained B_{max} values of between 0.6 and 3.2 fmol per mg protein and K_D values of between 30 and 84 pM for these patients. These values are similar to those previously found in other excitable membranes^{4,5,7}. Apamin receptors are completely absent from membranes of normal human muscle (Table 1; Fig. 1*b*).

Detailed studies of apamin receptors and of apamin action on Ca^{2+} -activated K^+ channels have been performed previously^{4–7}. Neither apamin receptors nor apamin-sensitive Ca^{2+} -activated K^+ channels are present in adult rat muscle⁵. However, both apamin receptors and apamin-sensitive Ca^{2+} -activated K^+ channels are present: (1) in non-innervated rat myotubes in culture^{4–6}; (2) *in vivo* during the embryonic period before the final pattern of innervation has been reached⁵; and (3) after denervation of adult rat muscle⁵. These observations prompted us to measure levels of apamin receptors in another series of patients suffering from spinal anterior horn disorders. None of them (Table 1) had any detectable apamin receptors on muscle membranes, which suggests that, if apamin receptors are present in these biopsies, they are below the detection level (0.3 fmol per mg protein). A simple explanation is that membranes belonging to totally denervated muscle are a very small proportion of

the total membrane preparation. Such a situation is likely to occur because muscle biopsies of patients suffering from spinal anterior horn disorders contain a mixture of normally innervated fibres and reinnervated fibres with polyinnervation; there is only a small proportion (<20%) of completely denervated fibres. The atrophied character of denervated fibres is another factor that reduces the chance of detecting apamin receptors.

Various other defects of myotonic dystrophic muscle membranes have previously been suggested, such as changes in levels and activity of the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ ⁸ or changes in insulin sensitivity⁹. The interest of the findings presented here is in the all-or-none character of the presence of the apamin receptor. It is detected in myotonic muscular dystrophy and is absent in normal human muscle or in muscles of patients suffering from

Table 1 ^{125}I -apamin-binding characteristics

	K_D (pM)	B_{max} (fmol per mg)
Patients with myotonic muscular dystrophy (11)	53 ± 6	1.5 ± 0.3
Patients with spinal anterior horn disorder (11)	Undetectable binding	
Patients taken as normal (10)	Undetectable binding	

B_{max} and K_D values of ^{125}I -apamin–receptor complex in myotonic muscular dystrophy are derived from Scatchard plots. The sample included 11 patients (7 male, 4 female, aged 17–69 yr). All showed characteristic clinical features and had myotonic discharges on EMG examination. The biopsied muscles were quadriceps (5 cases), peroneus longus (3), flexor carpi radialis (1), biceps (1) and deltoid (1). In each case the histological and histochemical appearance was consistent with the diagnosis. Three out of 11 cases showed no detectable binding. Eleven patients with disorders of the spinal anterior cells were studied (6 with amyotrophic lateral sclerosis, 1 with progressive muscular atrophy and 4 with spinal muscular atrophy; 7 male, 4 female, aged 22–70 yr). The diagnoses were confirmed by characteristic clinical findings. Nerve conduction was normal in all cases and EMG studies showed evidence of denervation and reinnervation. Histological and histochemical findings were consistent with a diagnosis of denervating disorders in all cases. Apamin binding was studied in quadriceps (8 cases), biceps (1), deltoid (1) and gastrocnemius (1). In no case was detectable binding observed. Ten patients (4 male, 6 female, aged 16–57 yr) undergoing surgery and showing no specific clinical features were categorized as normal. In no case was detectable binding observed. Numbers in parentheses indicate the number of patients.

spinal anterior horn disorders. Therefore, the receptor seems to be a good membrane marker for myotonic muscular dystrophy in membranes of muscle cells.

It remains to establish an association between the appearance of apamin receptors in the muscles of myotonic muscular dystrophy patients and the electrophysiological properties of biopsied muscle fibres. The most important property of myotonic dystrophic muscle fibres is their tendency to fire repetitive action potentials^{3,10,11}, which is related to EMG observations indicating that myotonic muscular dystrophy is characterized by myotonic runs³, that is, by repetitive bursts of activity. Another electrophysiological feature of the myotonic dystrophic muscle membrane is that it has a less negative resting potential¹²⁻¹⁵. One way to generate repetitive action potentials is the combination of a decreased resting potential with a value close to the threshold for the activation of Na⁺ channel and the presence of apamin-sensitive Ca²⁺-activated K⁺ channels, creating an after-hyperpolarization following the action potential which permits the transient reactivation of the Na⁺ channel after activation and inactivation during the action potential. It is the combination of such effects that creates repetitive bursts of activity responsible for spontaneous contractions in rat muscle cells in culture¹⁶. The same combination seems to be present in myotonic muscular dystrophy.

We thank Drs J. Pouget, J. F. Pellissier, B. Kullmann and P. Martin for biopsies and clinical studies, C. Roulinat-Bettelheim for technical assistance, and N. Chaudron for careful reading of the manuscript. This work was supported by the Associations Françaises de Lutte contre les Myopathies, the Centre National de la Recherche Scientifique (ATP 381) and the Ministère de l'Industrie et de la Recherche (grant 83.C.0918).

Received 28 October 1985; accepted 7 January 1986.

1. Grimm, T. *J. hum. Genet.* **23**, 301-308 (1975).
2. Todorov, A., Jequier, M., Klein, D. & Morton, N. E. *J. hum. Genet.* **18**, 387-406 (1970).
3. Rüdel, R. & Lehmann-Horn, F. *Physiol. Rev.* **65**, 310-356 (1985).
4. Hugues, M. *et al. EMBO J.* **9**, 1039-1042 (1982).
5. Schmid-Antomarchi, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2188-2191 (1985).
6. Romey, G. & Lazdunski, M. *Biochem. biophys. Res. Commun.* **118**, 669-674 (1984).
7. Hugues, M., Romey, G., Duval, D., Vincent, J. P. & Lazdunski, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1308-1312 (1982).
8. Desnuelle, C., Lombet, A., Serratrice, G. & Lazdunski, M. *J. clin. Invest.* **69**, 358-367 (1982).
9. Moxley, R. B., Corbett, A. J., Minaker, K. L. & Rowe, J. W. *Ann. Neurol.* **15**, 157-162 (1984).
10. Lipicky, R. J. in *Pathogenesis of Human Muscular Dystrophy* (ed. Rowland, L. P.) 729-739 (Excerpta Medica, Amsterdam, 1977).
11. Tahmouh, A. J., Askanas, V., Nelson, P. G. & Engel, W. K. *Neurology* **33**, 311-316 (1983).
12. McComas, A. J. & Mrozek, K. *J. neurol. Neurosurg. Psychiat.* **31**, 441-447 (1968).
13. Gruener, R., Stern, L. Z., Payne, C. & Hannapel, L. *J. neurol. Sci.* **24**, 339-349 (1975).
14. Gruener, R., Stern, L. Z., Markovitz, D. & Gerdes, C. *Muscle Nerve* **2**, 165-172 (1979).
15. Merickel, M., Gray, R., Chauvin, P. & Apple, S. *Proc. natn. Acad. Sci. U.S.A.* **78**, 648-652 (1981).
16. Barrett, E. F., Barrett, E. F. & Dribbin, L. B. *Dev. Biol.* **82**, 258-266 (1981).
17. Hartree, E. F. *Analyt. Biochem.* **48**, 422-427 (1972).

Inhibition of growth factor-induced differentiation of PC12 cells by microinjection of antibody to *ras* p21

Nabil Hagag, Simon Halegoua* & Michael Viola

Departments of Medicine and *Neurobiology and Behavior, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

The protein products (p21) of the *ras* cellular proto-oncogenes are thought to transduce membrane signals necessary for the induction of cell division¹⁻⁴. However, there is uncertainty as to the precise role of *ras* p21 in mediating ligand-membrane receptor signals leading to cell differentiation. Treatment of rat pheochromocytoma cells (PC12) with nerve growth factor (NGF) results in the induction of a number of phenotypic characteristics of sympathetic neurones, including cessation of cell division and outgrowth of neuronal processes (neurites). Here we report that microinjection of antibody to *ras* p21 into PC12 cells inhibited neurite formation and resulted in temporary regression of partially extended neurites, an effect which was observed up to 36 h after initiation of NGF treatment. Neurite formation induced by cyclic AMP was unaffected by injection of anti-p21 antibody. These results indicate that p21 is involved in the initiation phase of NGF-induced neurite formation in PC12 cells and has a role in hormone-mediated cellular responses distinct from cell proliferation.

PC12 cells are small (10-20 µm diameter) and poorly adherent to the plate surface and thus have a low viability after micropipette impalement. Therefore, to facilitate the present study, we fused PC12 cells with polyethylene glycol⁵ (see Fig. 1 legend). The resulting large multinucleated cells (~100 µm diameter) are responsive to NGF and express the same neuronal properties as unfused cells⁵. Treatment of fused cells with NGF (2.5S, 100 ng ml⁻¹) results in the formation of small processes by 12 h and these extend to long thin neurites which invariably terminate in growth cones by 36 h in >90% of cells. This is a faster time course than is observed in unfused cells⁶. Rat monoclonal anti-*ras* p21 antibody Y13-259 (ref. 7) was injected into PC12 cells immediately before NGF treatment. This antibody has been

shown to block both serum-induced growth of quiescent 3T3 cells and cell division induced by microinjection of *ras* p21 (ref. 2). The morphological differentiation of PC12 cells was scored by determining the percentage of cells having neurite processes longer than one cell body diameter and displaying a growth cone at their tip. In eight separate experiments (>500 cells injected), microinjection of antibody 259 at a concentration of 8 mg ml⁻¹ inhibited ~85% of injected cells from forming neurites (Fig. 1). At 8 mg ml⁻¹, 3.6 × 10⁶ IgG molecules would be injected per cell. Antibody at a concentration of 2 mg ml⁻¹ inhibited 64-76% of cells injected. This inhibitory effect was not seen in control cells after microinjection of either an isotype-identical irrelevant mouse monoclonal antibody or bovine serum albumin at the same concentration. The inhibitory effect of anti-p21 antibody lasted for ~48 h, after which neurite outgrowth was observed. After 3 days, the cultures injected with anti-p21 antibody had long neurites of normal appearance. The transient effect of anti-p21 antibody is consistent with previous studies which have shown degradation of microinjected protein molecules by 48 h post-injection⁸.

To determine how long after initiation of NGF treatment anti-p21 antibody can induce an inhibitory effect, a time course analysis was conducted. Fused PC12 cells were treated with NGF and different groups of cells were injected with antibody at 6-h intervals. Twenty cells were injected at each time point. Anti-p21 antibody inhibited neurite formation up to 36 h after initiation of NGF treatment, after which time no effect was noted. The percentage of cells inhibited at each time period was in the range 62-100%, with a mean of 82%. Injection of PC12 cells after neurite formation had begun (18-24 h) resulted in cessation and often retraction of neurite outgrowth. Neurite retraction began within 2-12 h and continued for 24 h, after which neurite outgrowth resumed (Fig. 2). During this period injected cells had a normal, healthy appearance. Thus, there appears to be a critical period in the initiation of the NGF effect in which PC12 cells are sensitive to the inhibitory effects of anti-p21 antibody.

Treatment of fused PC12 cells with dibutyryl cAMP resulted in neurite formation, as has been demonstrated in unfused PC12 cells⁹. However, neurite formation induced by dibutyryl cAMP was more rapid and the neurites were not as fine and extended as those seen after NGF treatment (Fig. 1). Approximately 200 cells were injected with anti-p21 antibody and a similar number

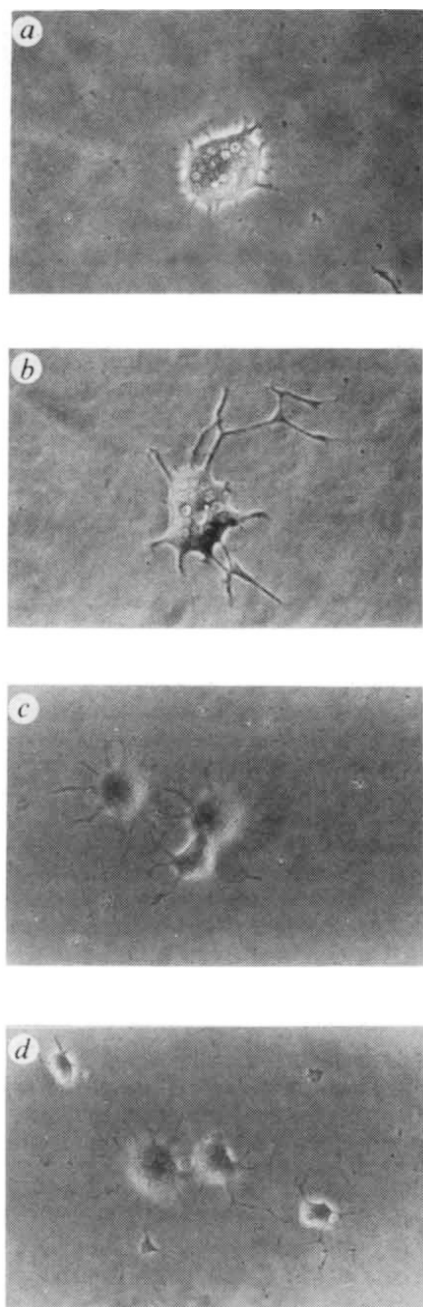


Fig. 1 Effects of microinjection of anti-p21 antibody on NGF- and cAMP-induced differentiation of PC12 cells. The figure shows phase-contrast micrographs of fused PC12 cells: *a*, injected with anti-p21 antibody 259 (8.0 mg ml^{-1}) followed by treatment with NGF (100 ng ml^{-1}) for 20 h; and *b*, injected with control antibody followed by NGF treatment as in *a*. Fused cells were also injected with anti-p21 antibody (*c*) and control antibody (*d*) then treated with dibutyryl cAMP (1 mM) for 20 h. Note that the anti-p21 antibody inhibits NGF- but not cAMP-induced neurite formation. $\times 100$.

Methods. PC12 cells were grown to confluence in Dulbecco's minimal essential medium (DMEM) supplemented with 10% fetal calf serum and 5% horse serum and incubated at 37°C in 10% CO_2 . Cells were fused with polyethylene glycol using the method of O'Laigue *et al.*⁵ and cells of the desired size ($100\text{--}200 \mu\text{m}$) were fractionated on a 5–30% serum gradient (to be published elsewhere), pooled, pelleted by centrifugation, resuspended in DMEM and plated on 35-mm plastic Petri dishes pre-coated with polylysine ($25 \mu\text{g ml}^{-1}$) and laminin ($10 \mu\text{g ml}^{-1}$). Fused cells were microinjected 24 h after plating. Antibodies were purified from hybridoma cells grown in serum-free medium and concentrated using Protein A-agarose (Bio-Rad) columns.

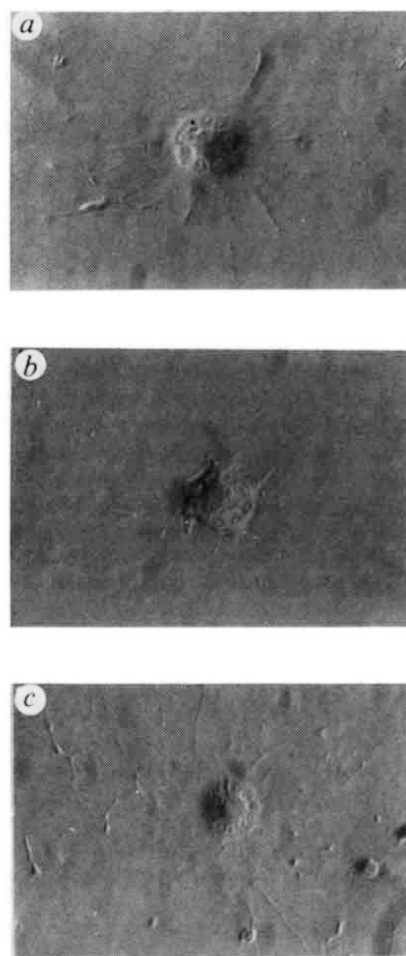


Fig. 2 Effect of microinjection of anti-p21 antibody on neurites induced in PC12 cells by NGF. The same cell is shown 36 h after the start of NGF (100 ng ml^{-1}) treatment, immediately before microinjection (*a*) and at 12 h (*b*) and 36 h (*c*) after injection of antibody 259 (8 mg ml^{-1}). NGF (100 ng ml^{-1}) treatment was maintained for the duration of the experiment. Note that neurites had almost completely retracted by 12 h after microinjection but had resumed normal neurite outgrowth with growth cone formation by 36 h post-injection. $\times 100$.

were injected with control antibody, immediately before treatment with dibutyryl cAMP; there was no difference between the two groups in either rapidity of neurite outgrowth, length of neurites or neurite morphology. These results demonstrate that anti-p21 antibody does not interfere with neurite formation *per se* but specifically inhibits neurite formation induced by NGF.

Treatment of PC12 cells with NGF results in the expression of several properties characteristic of sympathetic neurones, including cessation of cell division, development of long neurites, biosynthesis and storage of neurotransmitters and development of a Na^+ -based action potential¹⁰. The molecular mechanisms mediating NGF-induced phenotypic changes in PC12 cells have not been clearly defined, although multiple pathways appear to be involved. This proposal is based on the observations that NGF induces a rapid increase in cAMP levels¹¹ and stimulates phosphoinositide turnover¹² in PC12 cells. Further, one of us (S.H.) has shown that NGF induces the phosphorylation and activation of tyrosine hydroxylase at two sites, one mediated by cAMP-dependent protein kinase and the other by Ca^{2+} /phospholipid-dependent protein kinase (protein kinase C)^{13,14}. *ras* p21 has similar biochemical properties^{15–17} and shows sequence homology^{4,18,19} to GTP-binding proteins (G-proteins), molecules which regulate adenylate cyclase activity.

Further, the finding of activation of yeast adenylate cyclase by human and yeast *ras* gene products²⁰ has led to speculation that *ras* p21 mediates its effects in mammalian cells by altering the concentration of cAMP, although a recent study²¹ has been unable to demonstrate activation of mammalian adenylate cyclase by p21. The inhibition of NGF-induced neurite formation by anti-p21 antibody suggests that p21 mediates its effects through the polyphosphoinositide and/or cAMP-dependent systems and that NGF-induced PC12 cell differentiation may be a useful system to delineate precisely the metabolic pathways responsive to *ras* p21.

In the present study we have used a rat monoclonal antibody (Y13-259) generated against Harvey *v-ras* p21 which recognizes all rat p21 species⁷. Antibody 259 has been used extensively to identify and isolate products of both endogenous and transfected *ras* genes in a number of distantly related species^{3,17,22,23}. In previous microinjection experiments, antibody 259 was used to identify a requirement for *ras* p21 in serum-induced growth stimulation of rodent fibroblast cells². Beckner *et al.*²¹ have recently demonstrated a low-affinity, weak interaction of antibody 259 with N_i and N_s proteins (guanine nucleotide regulatory proteins which mediate inhibition and stimulation, respectively, of adenylate cyclase), in contrast to the high-affinity binding of this antibody with p21. Nevertheless, a remote possibility exists that antibody 259 has slight cross-reactivity with known and as yet unidentified GTP-binding proteins, resulting in the inhibition of the NGF response in PC12 cells.

After treatment of PC12 cells with NGF, the concentrations of Kirsten *c-ras* (*c-ras*^{Ki}) and Harvey *c-ras* (*c-ras*^{Ha}) messenger RNAs increase three- to fourfold²⁴. Further, it has recently been shown that infection of PC12 cells with *ras*-containing sarcoma viruses²⁵ and microinjection of purified *ras* oncogene p21 (ref. 26) results in neurite formation in PC12 cells. We have microinjected purified unaltered *c-ras*^{Ha} and activated *c-ras*^{Ha} p21 (ref. 27) into single and fused PC12 cells. With activated p21 the initial phase of differentiation of PC12 cells was noted—cell flattening and short neurite formation—but this was not as extensive as that seen after NGF treatment (N.H., S.H., M.V. and J. C. Lacal, unpublished observations). We conclude from the results of the experiments reported here that *ras* p21 function is necessary for NGF-induced neurite formation in PC12 cells and that the hormone-induced phenotypic changes mediated by *ras* p21 are an inherent property of the specific responsive cell type.

Expression of a new tyrosine protein kinase is stimulated by retrovirus promoter insertion

Anna F. Voronova*† & Bartholomew M. Sefton*

* Molecular Biology and Virology Laboratory, The Salk Institute, Box 85800, San Diego, California 92138, USA

† Department of Biology, University of California at San Diego, La Jolla, California 92093, USA

Tyrosine protein kinases are important both in the normal regulation of cellular proliferation and in the oncogenic transformation of cells by several tumour viruses¹. The LSTRA Moloney murine leukaemia virus (M-MuLV)-induced thymoma cell line² contains ~20-fold more phosphotyrosine in protein than do typical haematopoietic cell lines³; this seems to result from the expression of an abnormally high level of a cellular tyrosine protein kinase termed p56^{lck} (refs 3, 4). This kinase is normally expressed at low levels in most, but not all, murine T cells^{3,4}. The elevated levels of p56^{lck} could contribute to the malignant properties of LSTRA cells. Therefore, we have isolated cloned complementary DNAs encoding the whole of p56^{lck}. Sequence analysis shows it to be a novel cellular tyrosine protein kinase which is distinct from all others described to date. p56^{lck} is encoded in LSTRA cells by a hybrid messenger RNA; ~200 nucleotides at the 5' end of the mRNA are identical to the 5' end of the genome of M-MuLV. The three- to ninefold transcriptional activation of the gene therefore results from retroviral promoter insertion.

A major site of tyrosine phosphorylation in p56 is present in a tryptic peptide identical to that containing a phosphorylated tyrosine in p60^{v-src} (ref. 3), the transforming protein of Rous sarcoma virus. We therefore used a mixture of 32 synthetic heptadecanucleotides containing all the possible sequences encoding a portion of this peptide (Fig. 1a) as a probe to identify cDNA clones derived from the mRNA for p56 from LSTRA cells (Fig. 1b). Eighteen of 22 positive clones were related to each other and homologous with an ~2-kilobase (kb) mRNA expressed at an elevated level in LSTRA cells (see below). This mRNA contained a single open reading frame beginning at the first AUG at position 194, and encoded a protein of 509 amino acids (Fig. 1c). The predicted relative molecular mass of this polypeptide, excluding the initiator methionine, is 57,812.

The polypeptide contains a presumptive site of *in vitro* tyrosine phosphorylation which is identical in sequence to that in p56 (Fig. 1). In addition it has extensive amino-acid sequence homology with all the known tyrosine protein kinases (Fig. 2). It is also encoded by a mRNA which is expressed at an elevated level in two cell lines containing high levels of p56, and not expressed in cells which do not contain p56 (Fig. 3). We conclude that these clones are derived from the mRNA for p56.

p56 has most extensive homology with p60^{c-src} (ref. 5), p90^{v-yes} (ref. 6) and p70^{v-grb} (ref. 7). The homology with p60^{c-src} (ref. 5) is particularly striking (Fig. 2). The two proteins are very similar in size and contain their catalytic domains at very similar positions near their carboxy termini (Fig. 2). The homology extends well beyond the catalytic domains. The regions between positions 68 and 231 in p56 and 88 and 253 in p60^{c-src}, all of which is outside the catalytic domain, have 50% sequence identity.

p60^{c-src} is phosphorylated at Ser17 by the cyclic AMP-dependent protein kinase⁸ and can be phosphorylated at Ser 12 and Ser 48 by protein kinase C^{9,10}. Significantly, p56 contains no serine residues at homologous positions near its amino terminus. This emphasizes the fact that these two proteins have no homology in their amino-terminal domains. Notably, p56 contains a tyrosine residue in a position identical to that of Tyr 527 of p60^{c-src}, a site of *in vivo* tyrosine phosphorylation in p60^{c-src} (ref. 11). Tyr 505 may therefore be phosphorylated in p56.

Received 15 October; accepted 16 December 1985.

1. Feramisco, J., Gross, M., Kamata, T., Rosenberg, M. & Sweet, R. *Cell* **38**, 109–117 (1984).
2. Mulcahy, T., Smith, M. & Stacy, D. *Nature* **313**, 241–248 (1985).
3. Kamata, T. & Feramisco, J. *Nature* **310**, 147–150 (1984).
4. Hurley, J., Simon, M., Teplow, D., Robishaw, J. & Gilman, A. *Science* **226**, 860–862 (1984).
5. O'Lague, P. & Huttner, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1701–1705 (1980).
6. Greene, T. & Tischler, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2424–2428 (1976).
7. Furth, M. E., Davis, L. J., Fleurdelys, B. & Scolnick, E. M. *J. Virol.* **43**, 294–304 (1982).
8. Stacy, D. W. & Allfrey, V. G. *J. Cell Biol.* **75**, 807–817 (1977).
9. Schubert, D., Heinemann, S. & Kidokoro, Y. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2579–2583 (1977).
10. Greene, L. A. & Shooter, E. M. *A. Rev. Neurosci.* **3**, 353–402 (1980).
11. Shubert, D. & Whitlock, C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4055–4058 (1977).
12. Traynor, A. E., Shubert, D. & Allen, W. R. *J. Neurochem.* **39**, 1677–1683 (1982).
13. Helegoua, S. & Patrick, J. *Cell* **22**, 571–581 (1980).
14. McTigue, M., Cremins, J. & Halegoua, S. *J. biol. Chem.* **260**, 9047–9056 (1985).
15. Shih, T. Y., Papageorge, A., Stokes, P. E., Weeks, M. O. & Scolnick, E. M. *Nature* **287**, 686–691 (1980).
16. Gibbs, J. B., Sigal, I. S., Poe, M. & Scolnick, E. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5704–5708 (1984).
17. Sweet, R. W. *et al. Nature* **311**, 273–275 (1984).
18. Gilman, A. G. *Cell* **36**, 577–579 (1984).
19. Lochrie, M. A., Hurley, J. M. & Simon, M. I. *Science* **228**, 96–99 (1985).
20. Broeck, D. *et al. Cell* **41**, 763–769 (1985).
21. Beckner, S., Hattori, S. & Shih, T. *Nature* **317**, 71–72 (1985).
22. Finkel, T., Der, C. & Cooper, G. *Cell* **37**, 151–158 (1984).
23. Pawson, T. *et al. Molec. cell. Biol.* **5**, 33–39 (1985).
24. Curran, T. R. & Morgan, J. *Science* **229**, 1265–1268 (1985).
25. Noda, M. *et al. Nature* **318**, 73–75 (1985).
26. Bar-Sagi, D. & Feramisco, J. *Cell* **42**, 841–848 (1985).
27. Lacal, J. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 5305–5309 (1984).

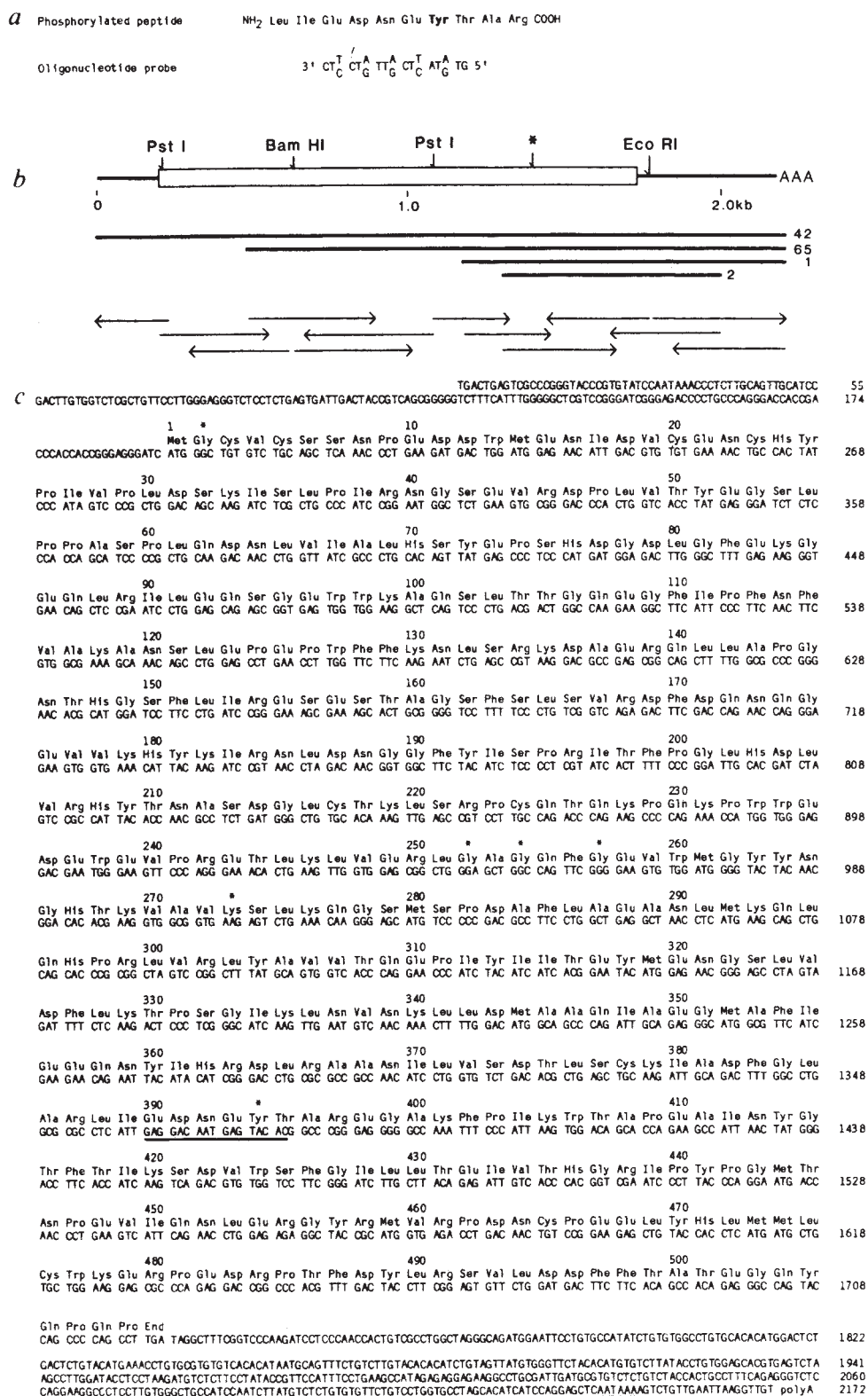


Fig. 1 Nucleotide sequence of mRNA and deduced amino-acid sequence of p56. **a**, Amino-acid sequence of the site of *in vitro* tyrosine phosphorylation in p56 and sequence of the mixture of oligonucleotides used as a probe. **b**, Schematic diagram of cDNA clones and mRNA encoding p56. Asterisk, position of the oligonucleotide probe; arrows, regions and strands subjected to sequence analysis. **c**, Nucleotide sequence and deduced amino-acid sequence of p56. The region complementary to the mixed probe (residues 1,361–1,377) and the polyadenylation signal (residues 2,148–2,151) are underlined. Comparison with the amino-acid sequences of the other members of the tyrosine protein kinase family suggests that the catalytic domain extends from residue 232 to 493. Asterisks: Gly 252, Gly 254, Gly 257 and Lys 272, which comprise a portion of the ATP-binding site³⁴, and Tyr 395, the site of *in vitro* phosphorylation.

Methods. The identity of the tryptic peptide containing the major site of *in vitro* tyrosine phosphorylation in p56 was confirmed by secondary digestion with *Staphylococcus aureus* V8 and *Pseudomonas fragii* proteases. The oligonucleotide probe was synthesized using phosphoramidite chemistry on a Systec Microsyn 1450 A oligonucleotide synthesizer. The cDNA was prepared from poly(A) mRNA isolated from LSTRA cells by standard procedures using reverse transcriptase and poly(dT)_{12–18} as a primer. The second strand was synthesized using RNase H and DNA polymerase I (ref. 35). The double-stranded cDNA was treated with S₁ nuclease. Homopolymeric dC tails were added with terminal transferase, and the cDNA was annealed to dG-tailed pBR322 (NEN). C600 *Escherichia coli* was transformed with the recombinant DNA using the Hanahan procedure³⁶. Approximately 80,000 recombinant colonies were screened by hybridization of replica nitrocellulose filters at 37 °C in 5×SSPE (1×SSPE is 0.18 M NaCl, 10 mM NaPO₄ pH 7.7, 1 mM EDTA), 2.5×Denhardt's, 100 µg ml⁻¹ homomix, 0.1% SDS, 0.5% NP40 (ref. 37) and washed in 2×SSPE, 0.1% SDS at 42 °C. Clones 1, 2, 42 and 65 were all subcloned in the mp19 derivative of M13 bacteriophage³⁸ using restriction endonucleases *Pst*I, *Bam*HI and *Eco*RI, and the nucleotide sequences were determined by the dideoxynucleotide chain termination method³⁹. A synthetic oligonucleotide primer was used to obtain the sequence of the 5' end of the cDNA.

Both p56 and p60^{c-src} are myristylated^{4,12,13}. In p60^{c-src}, the fatty acid is attached, through an amide bond, to the α-amino group of Gly 2, the amino-terminal residue of the mature protein^{14,15}. In every other case examined, myristic acid is also linked to an amino-terminal glycine¹². As codon 2 of the gene for p56 encodes glycine, we suggest that the initiating methionine residue is removed and that this glycine is the site of myristylation. The myristyl moiety in p60^{c-src} is crucial for the interaction of the protein with cellular membranes^{15,16}. It could play a similar role in p56.

p56 is present at a low level in most murine T cells, and at an elevated level in LSTRA and Thy 19 cells, but has not been detected in several other types of murine cells^{3,4,17}. A 2-kb transcript was detected in all T-cell lines examined but not in the RNA from a B-cell line or from fibroblasts (Fig. 3). LSTRA and Thy 19 cells were found to contain three- to ninefold more p56 mRNA than the other T cells examined. The abundance of mRNA for p56 is apparently not as elevated in these cells as is the abundance of the p56 polypeptide.

Both cell lines which express high levels of p56 are derived

p56	1	M G S S K S K P K D P S Q R R R S L E P P D S T H H G G F A S Q T P N I D V C	
p60 ^{c-src}	1	M G S S K S K P K D P S Q R R R S L E P P D S T H H G G F A S Q T P N I D V C	
	21	E N C H Y P I V L D S K I S L P I R N G S E V R D P L Y T Y E G S L P P A S P	
	41	P O T H R T P S R S F G T V A T E P K L F G G F N T S D T Y T S P Q R A G A L A	
	61	L Q D N L V I A I S Y E P S H D G D L G F E K G E Q L R T L E Q S G E W W K	
	81	G Y T T F V A L Y D Y E S R T E T D L S F E K K G E R L Q I V N N T E G W W K	
	100	A Q S L T T G Q E G F I P F N F A K A N S L E P E P F F K N L S R K D A E R	
	121	A H S L T T G Q E G F I P F N F A K A N S L E P E P F F K N L S R K D A E R	
	140	Q L L A P G N T H G S F L I R E S E T A G S F S L S V R D F D Q N G G E V V K	
	161	L L L A P G N T H G S F L I R E S E T A G S F S L S V R D F D Q N G G E V V K	
	180	H Y K I R N L D N G G F Y I S P R I T F P G G H D L V R H Y T N A S D G L C T K	
	201	H Y K I R K L D S G G F Y I T S R T Q F S L Q Q L V A Y Y S K H A D G L C H R	
	220	L S R P C Q T Q K P Q K P W W E D E W E V P R E T L K L Y E R L G A G Q F G E	
	241	L T N V C P T S K P Q T Q G L A K D A W E V P R E S L R L E V K L G Q G F G E	
	259	V W M G Y Y N G H T K V A V K S L K Q G S M S P D A F L A E A N L M K Q O H P	
	281	V W M G T W N G T T R V A I K L K P G N M S P E A F L O E A Q V M K L Q R H E	
	299	R L V R L Y A V Y T Q E P I Y I T E Y M E N G S L V D F L K T P S G I K L N V	
	321	K L V Q L Y A V Y S E E P I Y I T E Y M E N G S L V D F L K T P S G I K L N V	
	339	N K L L D M A A Q I A E G M A I I E Q N V Y H R D L R A A N I L V S D T L S C	
	361	P Q L V D M A A Q I A S G M A V E R M N Y V H R D L R A A N I L V S D T L S C	
	379	K I A D F G L A R L I E D N E Y T A R E G A K F P I K W T A P E A I N Y G T I F T	
	401	K V A D F G L A R L I E D N E Y T A R Q G A K F P I K W T A P E A A I Y G R F T	
	419	I K S D V M S F G I L L T E I Y T H G R I P Y P G M T N P E V I Q N L E R G Y R	
	441	I K S D V M S F G I L L T E I T T K G R V P Y P G M V N R E V L D Q Y E R G Y R	
	459	M Y R P D N C P E S L Y H L M C W K K E R P E D R P T F E Y L R S V I D D F F	
	481	M P C P E C P E S L H D L M C W K K E R P E D R P T F E Y L R S V I D D F F	
	499	T A T E G Q Y Q P Q P 509	
	521	T S T E P Q Y Q P G E N L 533	

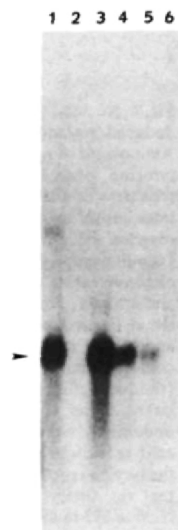
Fig. 2 Comparison of the deduced amino-acid sequence of p56 with that of p60^{c-src}. The amino-acid sequence of avian p60^{c-src} is from Takeya and Hanafusa⁵. The amino-acid sequence of p56 was deduced from the nucleotide sequence of clone 42. Identical amino acids are boxed.

from tumours induced by M-MuLV^{2,18}. Because avian leukaemia virus proviruses can activate transcription of the *c-myc* gene^{19,20} and the *c-erb-B* gene²¹ by promoter insertion, we asked whether the cDNA clones we had isolated from LSTRA cells had homology with M-MuLV genomic RNA. Strikingly, the 189 nucleotides at the 5' end of the cDNA clones were identical, apart from a single residue, to nucleotides 17–206 in the genome of M-MuLV (Fig. 4). The sequence of the cDNA diverged from that of the M-MuLV genome at the splice donor site which gives rise to the sub-genomic *env* mRNA²². This is five nucleotides upstream of the codon for the initiator methionine of p56.

Because the homology is almost complete, it is clear that the 5' portion of the p56 mRNA is derived from M-MuLV and not from an endogenous murine leukaemia virus. Transcription of the rearranged p56 gene in LSTRA cells is apparently initiated at the promoter present in the long terminal repeat (LTR) of an M-MuLV provirus. It is, therefore, simply a coincidence that the size of the mRNA for p56 in LSTRA cells is the same as in other T cells. The fact that the M-MuLV LTR contains an enhancer which is particularly active in T cells²³ may explain the elevated levels of p56 in LSTRA cells.

Retroviral proviruses contain LTRs at both of their termini, each of which contains a potentially active promoter²⁴. Transcription of the rearranged p56 gene in LSTRA cells must be initiated in a left-hand LTR because nucleotides downstream of only this LTR occur in the mRNA which gave rise to the cDNA clone. Two processes could yield a mRNA with the structure described here. First, a small fragment of an M-MuLV

Fig. 3 (Right) Agarose gel fractionation of mRNA. Lane 1, LSTRA cells; lane 2, S194 B cells; lane 3, Thy 19 cells; lane 4, Thy 16 cells; lane 5, BW 5147 cells; lane 6, Schmidt-Ruppin Rous sarcoma virus-transformed BALB 3T3 cells. All samples contained 5 µg of poly(A) mRNA, except 6, which contained 40 µg total cellular RNA. The samples were separated on a 1% agarose gel containing formaldehyde and transferred to nitrocellulose. Nick-translated clone 1 (Fig. 1) was used as a probe. Hybridization was performed at 42 °C in 50% formamide, 5 × SSPE, 2 × Denhardt's, 100 µg ml⁻¹ salmon sperm DNA, 0.1% SDS, 10% dextran sulphate. The filter was washed in 0.1 × SSPE, 0.1% SDS at room temperature. Comparison with the mobilities of 28S and 18S ribosomal RNA suggests a size of ~2 kb for the p56 mRNA.



provirus, consisting of only the left-hand LTR and 61 nucleotides of the untranslated L region, could be integrated immediately upstream from the coding portion of the p56 gene. Transcription originating at the viral promoter could then produce, without splicing, the mRNA we have observed. Alternatively, a provirus could be present anywhere upstream of the p56 gene and long transcripts arising from the 5' LTR could yield the p56 mRNA by splicing of the viral leader to a site just upstream from the coding region of the p56 gene.

We favour the second possibility. The sequence of the hybrid mRNA diverges from that of the M-MuLV genome at nucleotide 207 of the M-MuLV genome. This corresponds exactly to the splice donor site for the sub-genomic viral *env* mRNA²². This site could also function as a splice donor in the generation of mRNA for p56 in LSTRA cells. If so, this suggests that the normal nuclear precursor to the p56 mRNA contains a potential splice acceptor at nucleotide 189.

The reason for the over-expression of the p56 gene in Thy 19 cells is not yet understood. However, because both Thy 19 and LSTRA cell lines arose subsequent to infection with M-MuLV and because DNA blot analysis showed that both cell lines contain p56 genes which have undergone rearrangements at their 5' ends (data not shown), we suspect that the increased levels of p56 mRNA in Thy 19 cells also result from the integration of a M-MuLV provirus upstream of the gene.

Several sites of M-MuLV provirus integration have been characterized previously. These include the regions flanking the *c-myc*^{25,26}, *pim-1* (ref. 27), *mis-1/pvt-1* (refs 28, 29), *gin* (P. Jolicœur, personal communication), *Mlvi-1* (ref. 30) and *Mlvi-2* (ref. 31) genes. The p56 gene could conceivably be identical to *gin*, *Mlvi-1* (ref. 30) or *Mlvi-2* (ref. 31) genes. Hybridization experiments, however, have shown that it is not identical to either *pim-1* or *mis-1/pvt-1* (ref. 4).

The deduced amino-acid sequence of p56 shows it to be a new cellular tyrosine protein kinase. Although it has homology

p56	1	...TGACTGAGTCGCCGGGTACCCGTGATCCAATAAACCCCTCTTGCACTGCATCCGACTGTGGTCTCGCTGT	
M-MuLV	1	GCGCCAGTCCTCCGATTGACTGAGTCGCCGGGTACCCGTGATCCAATAAACCCCTCTTGCACTGCATCCGACTGTGGTCTCGCTGT	
	75	CCTTGGGAGGCTCTCTCTGAGTGATTGACTA CCGTCAGCGGGGGTCTTCATTGGGGCTCGTCCGGGATCGGGAGACCCCTGCCA	
	91	CCTTGGGAGGCTCTCTCTGAGTGATTGACTACCGTCAGCGGGGGTCTTCATTGGGGCTCGTCCGGGATCGGGAGACCCCTGCCA	
	164	GGGACCACCGACCCACCGAGGATCATGGGCTGTCTGCAGCTCAACCCCTGAAGATGACTGGATGGAGACCATGACGTGTGT	
	181	GGGACCACCGACCCACCGAGGATGAAGCTGGCCAGCACTTATCTGTGTCTGCCGATTGTCTAGTGTCTGACTGATTATTGCG	

Fig. 4 Comparison of the nucleotide sequence of the p56 mRNA from LSTRA cells with the nucleotide sequence of the genome of M-MuLV. The nucleotide sequence of M-MuLV is adopted from Shinnik *et al.*⁴⁰. Identical nucleotides are indicated by vertical bars; SD designates the splice donor site for the sub-genomic *env* mRNA of M-MuLV²². The position of the R and U5 regions are shown.

with all of the known tyrosine protein kinases, it also differs considerably in sequence from all of those described so far. In particular, its 60 amino-terminal residues have no homology with any known tyrosine protein kinase. In addition, the differences between the sequence of the murine gene for p56, the sequences of the avian genes for c-src and c-yes, and the feline gene for c-fgr are too great to reflect only evolutionary divergence. Much of the sequence of human p60^{c-src} has been deduced from cloned DNA. In marked contrast to p56 from murine cells, which is only 61% homologous with chicken p60^{c-src} in this region, human p60^{c-src} (ref. 32) is 98% homologous with chicken p60^{c-src}. We suggest the name *tck*, an acronym for T-cell kinase, for the gene encoding p56.

Viral tyrosine protein kinases are potent inducers of cellular transformation¹. Therefore, it is possible that the expression of p56^{tck} at an elevated level plays a role in the transformation of LSTRA and Thy 19 cells. As yet, however, there is no direct proof that the gene for p56^{tck} can function as an oncogene. With these cDNA clones in hand, we can now address this question.

We thank Willy Spaan for advice about the cloning of cDNA and nucleotide sequence determination; Chuck Van Beveren for critical discussions and help with the sequence comparisons; and Mark Kamps, Jan Buss and Sheila Podell for reading the manuscript. This work was supported by grants CA-14195 and CA-17289 from the US Public Health Service. A.F.V. was supported by training grant CA-09345 from the US Public Health Service.

Note added in proof: The *lsk*^T gene described recently by Marth *et al.*³³ is identical to the gene that we have termed *tck* here.

Received 24 September; accepted 5 December 1985.

An embryonic pattern of expression of a human fetal globin gene in transgenic mice

Kiran Chada, Jeanne Magram & Frank Costantini

Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, 701 W. 168th Street, New York, New York 10032, USA

During the evolution of the β -globin family gene in vertebrates, different globin genes acquired different developmental patterns of expression. In mammals, specific 'embryonic' β -like globins are synthesized in the earliest erythroid cells, which differentiate in the yolk sac of the embryo. In most mammals the embryonic globin chains are replaced by 'adult' β -globins in fetal and adult erythrocytes, which arise in the liver and bone marrow, respectively. However, in simian primates (including humans), a distinct 'fetal' type of β -like globin chain predominates in fetal erythroid cells^{1,2}. Based on the pattern of DNA sequence homologies between different mammalian species, these fetal globin genes, γ and $\Lambda\gamma$, are thought to have descended from an ancestral gene, 'proto- γ ', which was embryonic in its pattern of expression³⁻⁵. In the mouse, as well as in most other mammalian species, the descendants of the proto- γ gene continue to function as embryonic genes^{5,6}. To investigate the evolutionary changes that led to the 'fetal recruitment' of the γ -globin genes in primates, we have introduced the cloned human γ -globin gene into the mouse germ line. We report here that the human γ gene reverts to an embryonic pattern of expression in the developing mouse. This observation suggests that during evolution a shift occurred in the timing of expression of a *trans*-acting signal controlling the proto- γ gene.

The organization and evolutionary origin of the mouse and human β -globin gene clusters are shown in Fig. 1. A comparison of the nucleotide sequences of various mammalian globin genes has suggested that the last common ancestor to mouse and man contained five linked genes encoding β -like globin chains. This five-gene cluster evolved, by various processes including gene duplication and subsequent divergence, to yield the present-day gene clusters³⁻¹². In the lineage leading to simian primates^{8,9}, the descendant of the proto- γ gene ceased to be active in the earliest erythroid cells that differentiate in the yolk sac blood islands, and came to be expressed later in development, primarily in the fetal liver^{2,13}. There are several different mechanisms by which the pattern of γ -globin gene expression could have been altered during evolution; two simple models are presented in Fig. 2. Both models assume that the ancestral proto- γ gene was activated by a *trans*-acting signal present specifically in embryonic, but not fetal erythroid cells. In model I, a *cis*-acting regulatory sequence has been altered in the simian primate lineage, so that the human γ -globin gene is now recognized by a signal present in fetal erythroid cells (F), while the orthologous mouse gene continues to be recognized by an embryonic signal (E). Models of this type, involving changes in *cis*-acting regulatory elements, have been previously suggested to explain the fetal recruitment of the γ -globin genes^{5,9}. In model II, the γ -globin gene continues to be recognized by the same signal (γ) in both mouse and human, but in the human this signal is no longer produced in early embryonic erythroid cells, and it appears instead in fetal erythroid cells. For simplicity, the fetal to adult haemoglobin switch, which must involve additional regulatory mechanisms, is not considered here.

The ability to introduce cloned genes into the mouse germ line¹⁴⁻¹⁷ provided a means of experimentally distinguishing between these two types of evolutionary mechanisms. If a human γ -globin gene had lost the ability to recognize an embryonic signal and had evolved to respond instead to a signal in fetal erythroid cells, the gene might be expressed in fetal but not

- Sefton, B. M. & Hunter, T. *Adv. Cyclic Nucleotide Protein Phosphorylation Res.* **18**, 195-226 (1984).
- Glynn, J. P., Bianco, A. R. & Goldin, A. *Cancer Res.* **24**, 502-508 (1964).
- Casnellie, J. E., Harrison, M. L., Hellstrom, K. E. & Krebs, E. G. *J. biol. Chem.* **258**, 10738-10742 (1983).
- Voronova, A. F., Buss, J. E., Patschinsky, T., Hunter, T. & Sefton, B. M. *Molec. cell. Biol.* **4**, 2705-2713 (1985).
- Takeya, T. & Hanafusa, H. *Cell* **32**, 881-890 (1983).
- Kitamura, N., Kitamura, A., Toyoshima, K., Hirayama, Y. & Yoshida, M. *Nature* **297**, 205-208 (1982).
- Naharro, G., Robbins, K. C. & Reddy, R. P. *Science* **223**, 63-66 (1984).
- Collett, M. S., Erikson, E. & Erikson, R. L. *J. Virol.* **29**, 770-781 (1979).
- Gould, K. L. *et al. Cell* **42**, 849-857 (1985).
- Purchio, A. F., Shoyab, M. & Gentry, L. E. *Science* **229**, 1393-1395 (1985).
- Cooper, J. A., Gould, K. L., Cartwright, C. A. & Hunter, T. *Science* (submitted).
- Buss, J. E. & Sefton, B. M. *J. Virol.* **53**, 7-12 (1985).
- Marchildon, G. A., Casnellie, J. E., Walsh, K. A. & Krebs, E. G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7679-7682 (1984).
- Schulz, A. M., Henderson, L. E., Oroszlan, S., Garber, E. A. & Hanafusa, H. *Science* **227**, 427-429 (1985).
- Kamps, M. P., Buss, J. E. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4625-4628 (1985).
- Cross, F. R., Garber, E. A., Pellman, D. & Hanafusa, H. *Molec. cell. Biol.* **4**, 1834-1842 (1984).
- Casnellie, J. E., Gentry, L. E., Rohrschneider, L. R. & Krebs, E. G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6676-6680 (1984).
- Vogt, M. *Virology* **93**, 226-236 (1979).
- Hayward, W. S., Neel, B. G. & Astrin, S. M. *Nature* **290**, 475-480 (1981).
- Payne, G. S., Bishop, J. M. & Varmus, H. E. *Nature* **295**, 209-214 (1982).
- Nilsen, T. *et al. Cell* **41**, 719-726 (1985).
- Mann, R. & Baltimore, D. *J. Virol.* **54**, 401-407 (1985).
- Chatis, P. A., Holland, C. A., Hartley, J. W., Rowe, W. P. & Hopkins, N. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4408-4411 (1983).
- Varmus, H. E. *Science* **216**, 812-820 (1982).
- Steffen, D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2097-2103 (1984).
- Corcoran, L. M., Adams, J. M., Dunn, A. C. & Cory, S. *Cell* **37**, 113-122 (1984).
- Cuyper, H. T. *et al. Cell* **37**, 141-150 (1984).
- Lemay, G. & Jolicœur, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 38-42 (1984).
- Webb, E., Adams, J. M. & Cory, S. *Nature* **312**, 777-779 (1984).
- Tsichlis, P. N., Strauss, P. G. & Hu, L. F. *Nature* **302**, 445-449 (1983).
- Tsichlis, P. N., Strauss, P. G. & Kozak, C. R. & Cory, S. *Cell* **4**, 997-1000 (1984).
- Anderson, S. K., Gibbs, C. P., Tanaka, A., Kung, H. & Fugita, D. J. *Molec. cell. Biol.* **5**, 1122-1129 (1985).
- Marth, J. D., Peet, R., Krebs, E. G. & Perlmutter, R. M. *Cell* **43**, 393-404 (1985).
- Kamps, M. P., Taylor, S. S. & Sefton, B. M. *Nature* **310**, 589-592 (1984).
- Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
- Hanahan, D. J. *molec. Biol.* **166**, 557-580 (1983).
- Meinkoth, J. & Wahl, G. *Analyst. Biochem.* **138**, 267-284 (1984).
- Messing, J. *Meth. Enzymol.* **101**, 20-79 (1983).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. *Nature* **293**, 543-548 (1981).

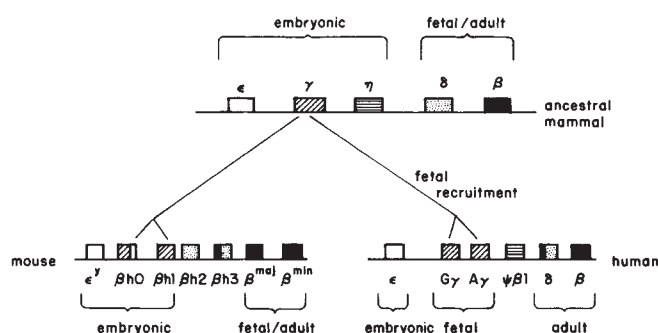


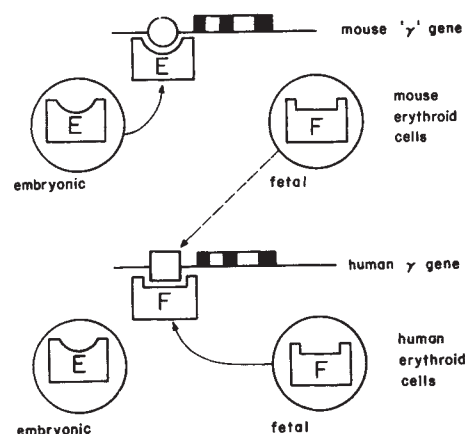
Fig. 1 Evolutionary origin of mouse and human β -globin gene clusters. The mouse cluster¹¹ includes three embryonic β -like globin genes (ϵ^y , $\beta H0$ and $\beta H1$) and two adult genes (β^{maj} and β^{min}); the human cluster¹² includes one embryonic gene (ϵ), two fetal genes ($G\gamma$ and $A\gamma$) and two adult genes (δ and β). These are thought to have evolved from a five-gene cluster in the last common ancestor^{3-5,7-9}. In each lineage, the proto- γ gene was independently duplicated, yielding the mouse genes $\beta H0$ and $\beta H1$ and the human genes $A\gamma$ and $G\gamma$ (refs 5, 22). The γ -derived genes remained embryonic in their expression in the mouse, while they were recruited for fetal expression in the human lineage. The evolution of a specialized fetal haemoglobin is thought to have fulfilled a physiological requirement for increased oxygen transport across the placenta¹. The derivations of the other mouse and human genes are indicated by their shading patterns^{4,7-9}, and the stage at which each is predominantly expressed is also indicated. The mouse genes $\beta H2$ and $\beta H3$, and the human gene $\psi\beta 1$ are pseudogenes^{11,41,42}.

embryonic erythroid cells in a transgenic mouse. On the other hand, if a human γ -globin gene were still regulated by the same signals that controlled the ancestral proto- γ gene, it might behave similarly to the orthologous mouse genes, $\beta H0$ and $\beta H1$, and revert to an embryonic pattern of expression in the mouse. Of course, *cis*-acting DNA sequences and *trans*-acting signals might have co-evolved since the divergence of rodents and primates, so that a human γ -globin gene in a developing mouse would fail to be expressed in a regulated pattern.

It has been reported previously that cloned adult β -globin genes display an appropriate pattern of developmental regulation in transgenic mice, remaining inactive in embryonic blood cells while being expressed in fetal and adult erythroid cells¹⁸⁻²⁰. These studies established that sequences close to the adult β -globin gene are sufficient to mediate its temporal activation during mouse development, and that an intact β -globin gene cluster is not a requirement for this step in haemoglobin switching. We therefore assumed, for the present experiment, that the *cis*-acting regulatory sequences of a γ -globin gene would also be closely linked to the gene.

We microinjected a 4-kilobase (kb) *Bgl*II-*Sph*I DNA fragment, containing the $G\gamma$ -globin gene^{21,22}, as well as 1.7 kb of 5' and 0.8 kb of 3' flanking DNA, into mouse zygotes^{23,24} and produced a series of transgenic animals. Six transgenic lines, carrying between 2 and 10 copies of the $G\gamma$ -globin gene (as determined by Southern blot analysis of tail DNA; data not shown), were derived and analysed for expression of the foreign gene. Total RNA was isolated from erythroid tissues at three different stages of development: (1) peripheral blood from adult mice, (2) livers from 14.5- or 15.5-day-old fetuses, and (3) circulating blood cells from 11.5-day-old embryos. The amount of human $G\gamma$ -globin messenger RNA in each RNA preparation was measured²⁵ (Fig. 3A). In three transgenic lines (two of which, G9 and G38, are shown), $G\gamma$ -globin mRNA was detected in the embryonic blood cells, but not in the fetal liver or adult blood cells. No $G\gamma$ -globin mRNA was detected at any stage in the remaining three lines; the lack of expression in a subset of transgenic lines has been observed for other globin genes, and has been discussed previously^{18,20}. In the one line (G9) that was tested for tissue specificity, expression of the $G\gamma$ -globin gene in the 11.5-day-old embryo appeared to be limited to blood cells

MODEL I



MODEL II

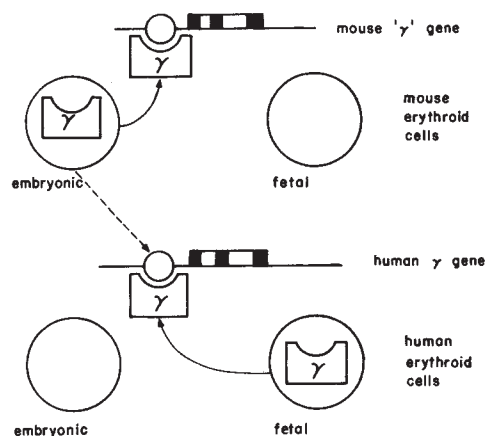
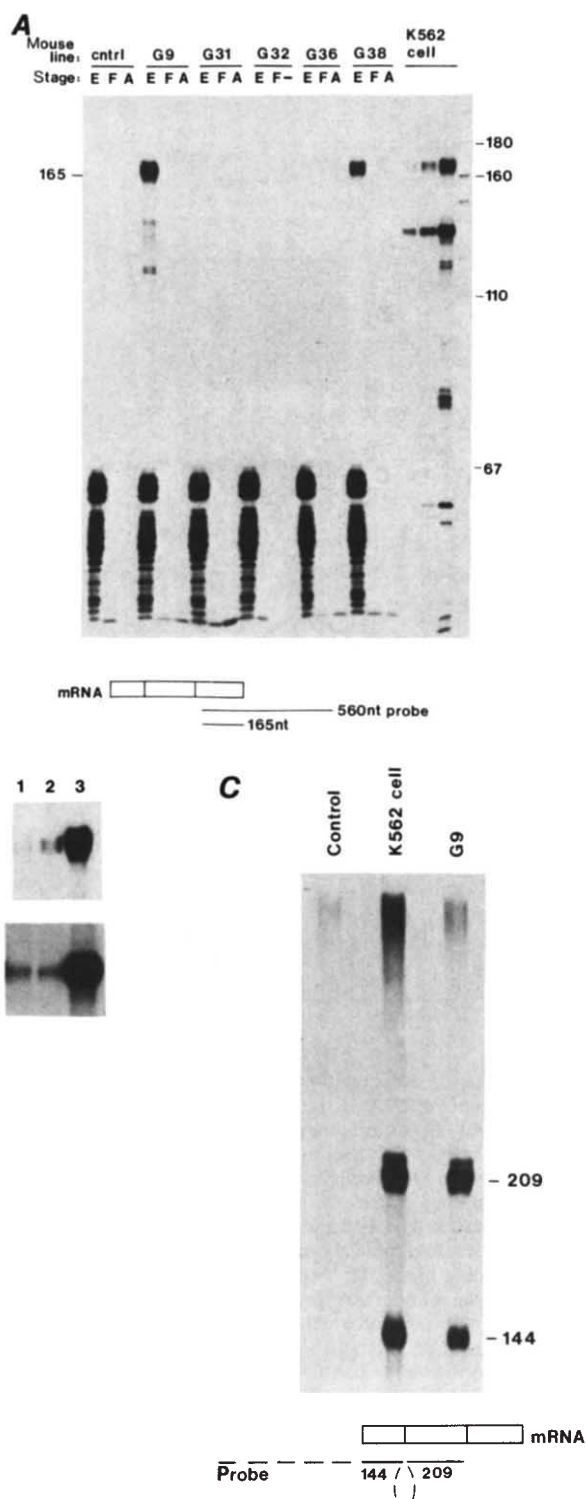


Fig. 2 Two models for the mechanism of fetal recruitment of the γ -globin gene in the human lineage. In model I, a γ -like gene in the mouse (as well as in the last common ancestor of mouse and man) is activated by a stage-specific factor, E, present in embryonic erythroid cells, which recognizes a *cis*-acting regulatory element (circle). In human, the *cis*-acting regulatory element (square) has been altered so that it is no longer recognized by factor E and is instead recognized by a different factor, F, present in fetal erythroid cells. In model II, a mouse (or ancestral) γ -like globin gene is activated by a gene-specific ' γ ' factor. In the mouse, this factor is present only in embryonic erythroid cells. In human, a γ -globin gene retains the same *cis*-acting regulatory element (circle), and is activated by the same factor, but the stage of erythroid development at which the factor appears has shifted from embryonic to fetal. The latter model requires the assumption that the proto- ϵ embryonic globin gene was regulated by a different signal from the proto- γ gene, since ϵ has remained an embryonic gene in primates as well as rodents. Consistent with this assumption, the γ -like embryonic genes in rodents, as well as in the distantly related mammalian order Lagomorpha, display a different timing of expression from the ϵ -like embryonic genes^{6,43}.

(Fig. 3B). The $G\gamma$ -globin mRNA was shown by S_1 nuclease analysis^{26,27} to be correctly initiated and spliced (Fig. 3C). Relative to endogenous $\beta H1$ mRNA, the levels of $G\gamma$ -globin mRNA in 11.5-day blood cells were 16% (line G9) and 4% (line G38), as determined by quantitative RNase protection measurements^{25,28} using RNA probes for the $\beta H1$ and $G\gamma$ mRNAs (data not shown). Embryonic blood cells synthesized a globin chain that co-migrated with human $G\gamma$ -globin chains on acid/urea/Triton polyacrylamide gels (data not shown) and which was precipitated by a monoclonal antibody²⁹ specific for human γ -globin chains (T. Papayannopoulou and G. Stamatoyannopoulos, personal communication).

Fig. 3 **A**, Stage-specific expression of the human γ -globin gene in mouse embryonic blood cells. Five founder transgenic male mice and a control mouse were mated with normal females, and total RNA was isolated^{19,44} from the circulating blood of a litter of 11.5-day-old embryos (E), and from the livers of 14.5- or 15.5-day-old fetuses (F). RNA was also isolated from peripheral blood of each adult mouse (A). Each RNA sample was hybridized with a ³²P-labelled RNA probe specific for human γ -globin mRNA, digested with RNase, then electrophoresed on a urea-polyacrylamide gel and an autoradiograph produced²⁵. The protected fragment of 165 nucleotides (nt) is due to specific hybridization to human γ -globin mRNA, while the products of low relative molecular mass obtained with all the mouse embryonic blood samples are due to cross-hybridization with endogenous embryonic globin mRNAs. The probe was synthesized from an *Eco*RI-*Hind*III fragment of the human γ -globin gene cloned in the vector pSP64 (ref. 25). The amount of each embryonic (or fetal) RNA sample used was $5/x \mu\text{g}$, where x is the fraction of embryos in each litter that were transgenic, determined by DNA dot-blot analysis (x ranged from 2/9 to 7/9); $5 \mu\text{g}$ of each adult blood RNA was used. RNA (1, 2 and $10 \mu\text{g}$) from human K562 cells induced with $20 \mu\text{M}$ haemin was hybridized as a positive control. K562 is a human leukaemic cell line that can be induced to undergo erythroid differentiation, expressing embryonic and fetal haemoglobins^{45,46}. **B**, Tissue-specific expression of the γ -globin gene. Blood cells were collected from a litter of 11.5-day-old embryos in line G9, and total RNA was prepared from the blood (lane 3) and from the upper (lane 1) and lower (lane 2) halves of the exsanguinated carcasses. The RNAs were hybridized with RNA probes specific for the βh1 and γ -globin mRNAs (as described above, and see also Fig. 4 legend). Only a very small amount of γ -globin mRNA was detected in the carcass, and this appears to be largely due to contamination by blood as a comparable amount of βh1 mRNA was found in the carcass. **C**, γ -Globin mRNA in transgenic mice is correctly initiated and spliced. Total RNA ($10 \mu\text{g}$) from either control 11.5-day-old embryonic blood cells, induced K562 cells, or 11.5-day-old transgenic embryonic blood cells (line G9), was hybridized with a uniformly labelled single-stranded DNA probe²⁷, and the hybrids were digested with S_1 nuclease and analysed by electrophoresis on a urea-polyacrylamide gel²⁶. The origin of the 209-nt and 144-nt protected fragments is shown schematically at the bottom.



Our results indicate that during mouse development an introduced human γ -globin gene behaves as an embryonic gene, suggesting that the gene may respond to the same *trans*-acting signals that regulate one or both of the orthologous mouse genes, βh0 and βh1 . As described previously⁶, the βh0 and βh1 genes are maximally expressed at a somewhat earlier stage than the other embryonic gene, ϵ^γ , during the maturation of embryonic erythroid cells. We therefore thought it interesting to compare the time course of accumulation of the γ -globin mRNA with that of the mouse βh1 - and ϵ^γ -globin mRNAs. As shown in Fig. 4, the levels of both endogenous βh1 mRNA and human γ -globin mRNA reached a maximum on day 11.5, and declined substantially by day 13.5, although the βh1 mRNA declined more rapidly; in contrast, the level of ϵ^γ mRNA increased continuously from day 10.5 to day 12.5 and declined only slightly on day 13.5. Although the relative contributions of synthesis and turnover to the observed kinetics of mRNA accumulation are unknown, these data are consistent with the coordinate transcription of the βh1 and γ -globin genes during the embryonic phase of erythropoiesis.

The embryonic stage-specific expression of a human γ -globin gene in transgenic mice suggests that the *trans*-acting signals that once regulated the expression of the proto- γ gene, and the DNA sequences with which they interacted have been functionally conserved since the divergence of mouse and man. Thus, the change in the pattern of γ -globin gene expression in simian primates cannot easily be attributed to mutations in *cis*-acting DNA sequences which rendered the gene unresponsive to the ancestral embryonic signals (model I). Rather, our results suggest that the γ -globin gene continued to be regulated by the same *trans*-acting signals in both the primate and rodent

lineages, while the timing of these signals changed in primates (model II). The specific models that we have described are probably oversimplified, as globin gene transcription may be controlled by multiple factors, both positive and negative. However, the conclusion that fetal recruitment involved a change in the timing of a *trans*-acting signal appears to be largely independent of the details of the regulatory mechanism. The alteration of patterns of gene regulation is thought to have been a crucial process in the evolution of mammals³⁰⁻³², and the type of change which we propose here may be a generally important mechanism of evolution.

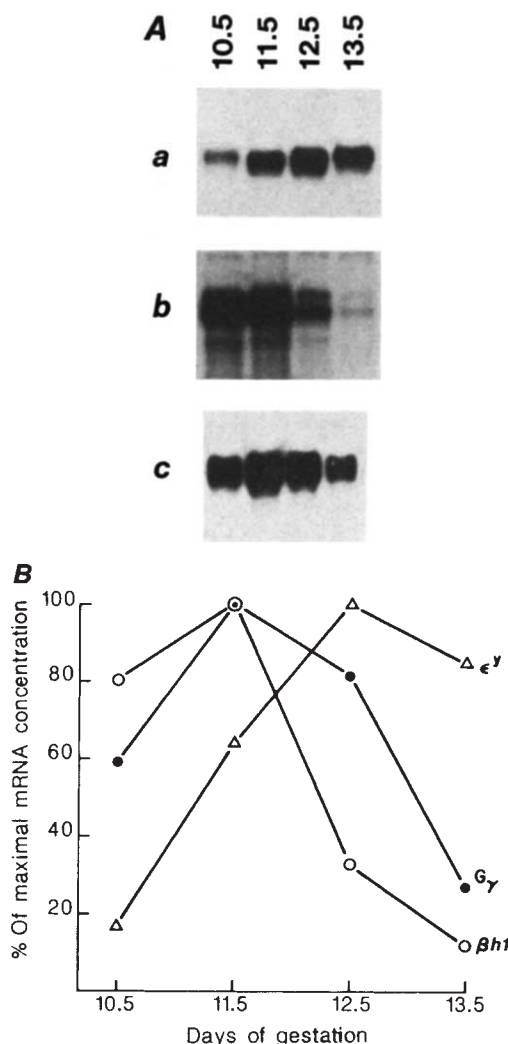


Fig. 4 Rate of accumulation of human G^γ and endogenous ϵ^γ and β^{h1} globin mRNAs in transgenic blood cells during embryonic development. **A**, Autoradiographic signals obtained in nuclease protection assays with probes specific for ϵ^γ (**a**), β^{h1} (**b**) and G^γ (**c**) mRNAs, in blood cells from transgenic embryos at days 10.5–13.5 of gestation. **B**, Relative change in the concentration of each mRNA between days 10.5 and 13.5, as determined by densitometry of the autoradiographic signals shown in **A**. Δ , ϵ^γ -Globin probe; $\circ$, β^{h1} globin probe; $\bullet$, G^γ -globin probe.

Methods. The founder transgenic male mouse (G9) was mated with normal females, one of which was killed on each of the indicated days of gestation (taking the morning on which the copulation plug was observed as day 0). The embryos were washed free of maternal blood in phosphate-buffered saline, and embryonic blood was collected from the entire litter. Total RNA was isolated⁴⁴ from the pooled blood cells of each litter, and DNA was isolated from each individual embryo. The number of embryos in each litter that inherited the human G^γ -globin gene was determined by dot-blot analysis of the DNAs. G^γ , ϵ^γ and β^{h1} mRNAs were measured by hybridization with specific DNA or RNA probes (in probe excess), followed by S_1 nuclease or RNase digestion, electrophoresis on urea-polyacrylamide gels, and autoradiography. The probes were: a uniformly labelled SP6 RNA probe for the 3' end of the G^γ -globin gene (see Fig. 3); an end-labelled probe for the 3' end of the ϵ^γ -globin mRNA¹⁹; and a uniformly labelled SP6 RNA probe for the 3' end of the β^{h1} globin mRNA (synthesized from a *Sau3A*–*Sau3A* fragment of the β^{h1} gene, cloned in the vector pSP65; see ref. 25). 20 ng of each RNA sample was used for hybridization with the ϵ^γ probe, and 1 μ g of each RNA sample was hybridized with the β^{h1} probe. Only a fraction of the embryos in each litter inherited the foreign G^γ -globin gene. Therefore, for hybridization with the G^γ -globin probe, the amount of RNA used was $5/x$ μ g, where x is the fraction of transgenic embryos in the litter (x ranged from 4/11 to 5/9).

Our studies also allow us to make several conclusions concerning the regulation of haemoglobin switching. A G^γ -globin gene with a limited amount of 5' and 3' flanking DNA appears to contain *cis*-acting regulatory sequences sufficient to dictate a stage-specific pattern of expression in the developing mouse. The region of DNA between the A^γ - and δ -globin genes, implicated as a possible negative regulatory region because its deletion is often associated with hereditary persistence of fetal haemoglobin (reviewed in ref. 33), was absent from the G^γ -globin gene we injected. Therefore, this region is not required for the gene to be switched off during the embryonic to fetal transition in mouse erythroid development. The embryonic pattern of expression displayed by the G^γ -globin gene is distinct from that of a cloned adult β -globin gene in the mouse germ line^{19,20}. Thus, it should be possible, using this experimental system, to define the regulatory DNA sequences responsible for the differential expression of embryonic and adult globin genes during mouse development.

When a human γ -globin gene is introduced, by DNA transfection or chromosome transfer, into murine erythroleukaemia (MEL) cells, which represent adult erythroid cells³⁴, the gene is usually expressed and in many cases its expression increases during induced MEL cell differentiation^{35–39}; this implies that the *trans*-acting factors responsible for inducible expression in murine adult erythroid cells are not specific for adult globin genes. In contrast, the stage-specific expression of the G^γ -globin gene in transgenic mice suggests the existence of additional regulatory factors that can distinguish between a γ - and a β -globin gene. These factors might control an earlier step in globin gene activation, one that is bypassed by genes introduced into cultured cells^{28,40} but not by genes present throughout development.

We thank R. Axel, R. Hardison, E. Lacy, T. Maniatis and M. McClelland for critical reading of the manuscript. K.C. was supported by a postdoctoral fellowship from NATO and F.C. by an Irma T. Hirsch Career Scientist Award. This work was supported by grants from the NIH (HD17704) and the March of Dimes (MDBDF 5-410).

Received 14 October; accepted 31 December 1985.

- Kitchen, H. & Brett, I. *Ann. N.Y. Acad. Sci.* **241**, 653–671 (1974).
- Weatherall, D. J. & Clegg, J. B. *The Thalassemia Syndromes* 3rd edn (Blackwell, Oxford, 1981).
- Czelusniak, J. *et al. Nature* **298**, 297–300 (1982).
- Hardison, R. C. *Molec. Biol. Evol.* **1**, 390–410 (1984).
- Hill, A. *et al. J. Biol. Chem.* **259**, 3739–3747 (1984).
- Farace, M. G. *et al. J. Biol. Chem.* **259**, 7123–7128 (1984).
- Hardies, S. C., Edgell, M. H. & Hutchison, C. A. III *J. Biol. Chem.* **259**, 3748–3756 (1984).
- Harris, S., Barrie, P. A., Weiss, M. L. & Jeffries, A. J. *J. molec. Biol.* **180**, 785–801 (1984).
- Goodman, M., Koop, B. F., Czelusniak, J. & Weiss, M. L. *J. molec. Biol.* **180**, 803–823 (1984).
- Efstratiadis, A. *et al. Cell* **21**, 653–668 (1980).
- Jahn, C. L. *et al. Cell* **21**, 159–168 (1980).
- Fritsch, E. F., Lawn, R. M. & Maniatis, T. *Cell* **19**, 959–972 (1980).
- Peschle, C. *et al. Nature* **313**, 235–238 (1985).
- Gordon, J., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. & Ruddle, F. H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7380–7384 (1980).
- Wagner, E. F., Stewart, T. A. & Mintz, B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5016–5020 (1981).
- Costantini, F. & Lacy, E. *Nature* **294**, 92–94 (1981).
- Brinster, R. L. *et al. Cell* **27**, 223–231 (1981).
- Chada, K. *et al. Nature* **314**, 377–380 (1985).
- Magram, J., Chada, K. & Costantini, F. *Nature* **315**, 338–340 (1985).
- Townes, T. M., Lingrel, J. B., Brinster, R. L. & Palmiter, R. D. *EMBO J.* **4**, 1715–1723 (1985).
- Slightom, J. L., Blechl, A. E. & Smithies, O. *Cell* **21**, 627–638 (1980).
- Shen, S.-H., Slightom, J. L. & Smithies, O. *Cell* **26**, 191–203 (1981).
- Costantini, F. & Lacy, E. *J. cell. Physiol. Suppl.* **1**, 219–226 (1982).
- Hogan, B., Costantini, F. & Lacy, E. *Manipulating the Mouse Embryo: Laboratory Manual* (Cold Spring Harbor Laboratory, New York, in the press).
- Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035–7056 (1984).
- Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
- Ley, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 6618–6622 (1984).
- Charnay, P. *et al. Cell* **38**, 251–263 (1984).
- Stamatoyannopoulos, G. *et al. Blood* **61**, 530–539 (1983).
- Britten, R. J. & Davidson, E. H. *Q. Rev. Biol.* **46**, 111–138 (1971).
- Wilson, A. C., Carlson, S. S. & White, T. J. *A. Rev. Biochem.* **46**, 573–639 (1977).
- Dobson, D. E., Prager, E. M. & Wilson, A. C. *J. Biol. Chem.* **259**, 11607–11616 (1984).
- Collins, F. S. & Weissman, S. M. *Prog. Nucleic Acid Res. molec. Biol.* **31**, 315–462 (1985).
- Marks, P. A. & Rifkind, R. A. *A. Rev. Biochem.* **47**, 419–426 (1978).
- Willing, M. C., Neinhuis, A. W. & Anderson, W. F. *Nature* **277**, 534–538 (1979).
- Wright, S., deBoer, E., Grosfeld, F. G. & Flavell, R. A. *Nature* **305**, 333–336 (1983).
- Papayannopoulou, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 780–784 (1985).

38. Anagnou, N. P., Karlsson, S., Moukton, A. D., Keller, G. & Neinhuis, A. W. *EMBO J.* (in the press).
39. Charnay, P. & Henry, L. *EMBO J.* (submitted).
40. Robins, D. M., Paek, I., Seeburg, P. H. & Axel, R. *Cell* **29**, 623-631 (1982).
41. Phillips, S. J., Hardies, S. C., Jahn, C. L., Edgell, M. H. & Hutchison, C. A. III *J. Biol. Chem.* **259**, 7947-7954 (1984).
42. Hutchison, C. A. III, Hardies, S. C., Padgett, R. W., Weaver, S. & Edgell, M. H. *J. Biol. Chem.* **259**, 12881-12889 (1984).
43. Rohrbaugh, M. L. & Hardison, R. C. *J. molec. Biol.* **164**, 395-417 (1983).
44. Chirgwin, J., Przybyla, A., MacDonald, R. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
45. Lozzio, C. B. & Lozzio, B. B. *Blood* **45**, 321-334 (1975).
46. Benz, E. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3509-3513 (1980).

Yeast plasma membrane ATPase is essential for growth and has homology with (Na⁺ + K⁺), K⁺- and Ca²⁺-ATPases

Ramón Serrano, Morten C. Kielland-Brandt & Gerald R. Fink

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

The plasma membrane ATPase of plants and fungi is a hydrogen ion pump¹. The proton gradient generated by the enzyme drives the active transport of nutrients by H⁺-symport. In addition, the external acidification in plants and the internal alkalization in fungi, both resulting from activation of the H⁺ pump, have been proposed to mediate growth responses. This ATPase has a relative molecular mass (*M_r*) similar to those of the Na⁺-, K⁺- and Ca²⁺-ATPases of animal cells and, like these proteins, forms an aspartylphosphate intermediate. We have cloned, mapped and sequenced the gene encoding the yeast plasma membrane ATPase (*PMA1*) and report here that it maps to chromosome VII adjacent to *LEU1*. The strong homology between the amino-acid sequence encoded by *PMA1* and those of (Na⁺ + K⁺), Na⁺-, K⁺- and Ca²⁺-ATPases is consistent with the notion that the family of cation pumps which form a phosphorylated intermediate evolved from a common ancestral ATPase. The function of the *PMA1* gene is essential because a null mutation is lethal in haploid cells.

The gene encoding the yeast plasma membrane ATPase was cloned by screening a λ gt11 expression library² with antibody prepared against the yeast enzyme. Antibody was obtained by injecting rabbits with purified protein that had been excised from a polyacrylamide gel^{3,4}. Seven positive clones were obtained by screening 3×10^6 plaques with antiserum that had been passed over a Sepharose column containing immobilized antigens from a λ gt11/*Escherichia coli* lysate to reduce the concentration of antibodies reacting with the bacterial vector and host². When these clones were rescreened with antibody produced by affinity chromatography on an ATPase-Sepharose column^{5,6}, only one of the seven clones was positive. The insert in the immunoreactive λ gt11 clone is an *EcoRI* fragment of 3.3 kilobases (kb) which, by Southern analysis, is present in the yeast genome in only a single copy. In Northern analysis, the insert hybridized with an RNA of 3.4 kb, which is large enough to encode the ATPase. The λ gt11 *EcoRI* fragment was cut into three fragments of roughly equal size: *EcoRI*-*SalI*, *SalI*-*XbaI* and *XbaI*-*EcoRI* (Fig. 1A, left to right). Each of these fragments was hybridized to total yeast RNA. The *EcoRI*-*SalI* fragment hybridized strongly to the 3.4-kb RNA, the *SalI*-*XbaI* fragment reacted only weakly with it and the *XbaI*-*EcoRI* fragment showed no hybridization. A Western blot^{7,8} of lysates prepared by thermo-induction of lysogens of the immunoreactive λ gt11 phage⁹ showed the presence of an antibody-binding protein of *M_r* ~170K. Based on the size of β -galactosidase, the fusion protein contains ~45K of the ATPase. This size estimate and that obtained by hybridizing subfragments of the insert to RNA suggest that the λ gt11 clone contains ~45% of the ATPase coding region at the carboxy-terminus of the enzyme. The orientation of the insert in λ gt11, determined by restriction mapping, is consistent with these data.

The insert from the positive λ gt11 clone was subcloned into the plasmid pUC18 (ref. 10) and used as a probe for colony hybridization of yeast genomic libraries in the plasmids YEp24 and YCp50. Four of 6,000 YEp24 colonies and three of 9,000 YCp50 colonies were positive. All seven clones contain the same 5-kb *HindIII* fragment that hybridizes with the λ gt11 insert. The restriction map of this fragment and some of the flanking sequences was determined after subcloning in pUC18 (see Fig. 1A). The ATPase gene was sequenced by the strategy described in the legend to Fig. 1B; the nucleotide sequence and the amino-acid sequence deduced from it are shown in Fig. 2. The

Fig. 1 A, Restriction map of the 5-kb *HindIII* fragment containing the ATPase gene and its flanking regions. The position of the λ gt11 insert, the ATPase coding region (with the direction of transcription), the *EcoRI* and *KpnI* fragments used for gene disruption, and the beginning of the *LEU1* coding region are indicated. B, Sequencing strategy. DNA sequencing was performed by the chemical cleavage method of Maxam and Gilbert²⁵, simplified as suggested by Bencini *et al.*²⁶. The 5-kb *HindIII* fragment was subcloned into pUC18 (ref. 10) and some sequence was obtained from the *HindIII*, *XhoI*, *EcoRI* and *SalI* sites (see A). Most of the sequence, however, was obtained by a linker insertion strategy²⁷. The *XhoI* site was first destroyed by cutting with *XhoI*, filling the 5' protruding ends with the Klenow fragment of DNA polymerase I and blunt-end ligation. The resulting plasmid was linearized by partial digestion with either *AluI*, *RsaI* or *TaqI* and the linker DNAs were purified by electroelution. *XhoI* linkers (GGAGCTCC; Collaborative Research) were ligated at the different cutting sites (in the case of *TaqI* cuts, after making blunt ends with Klenow enzyme). Sequencing was effected from the different linkers after cutting with *XhoI* and labelling with [α -³²P]dTTP (Amersham) and either Klenow enzyme (Boehringer) or reverse transcriptase (Life Sciences, St Petersburg, Florida). The fragments sequenced from each linker or restriction site are indicated.

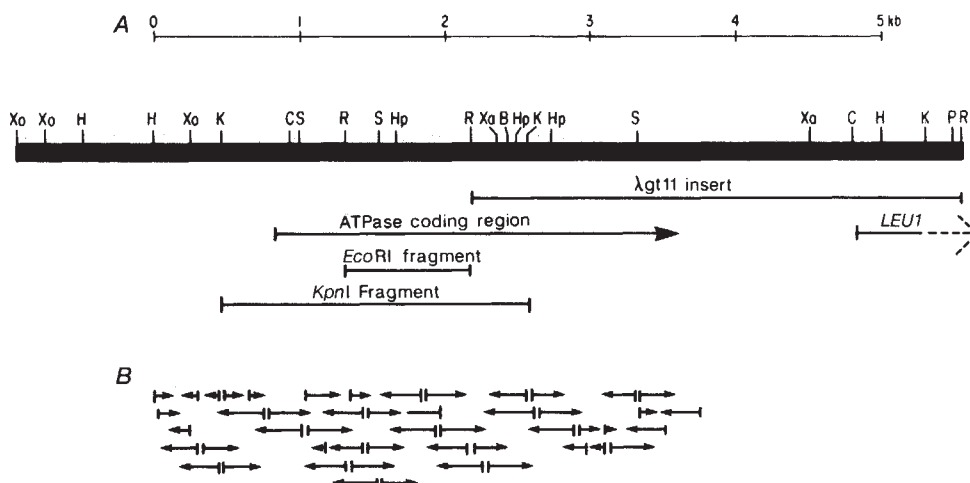


Fig. 2 Nucleotide sequence of the yeast plasma membrane ATPase gene, and the inferred amino-acid sequence. The nucleotide sequence starts at the *Xho*I site destroyed during the sequencing procedure (see legend to Fig. 1B). The poly(T)-rich region, the consensus TATA box and the upstream *Kpn*I site used for gene disruption (see Fig. 5) are indicated. The 10 hydrophobic stretches (see Fig. 4) are underlined.

Comparisons of the sequence of H⁺-ATPase with those of the K⁺-ATPase of *E. coli*¹², the (Na⁺+K⁺)ATPase from kidney¹³, and the Ca²⁺-ATPase from muscle¹⁴ show that these ATPases have extensive homologies even though they have very different transport capabilities (Fig. 3). The regions conserved in the four ATPases may correspond to portions of the proteins

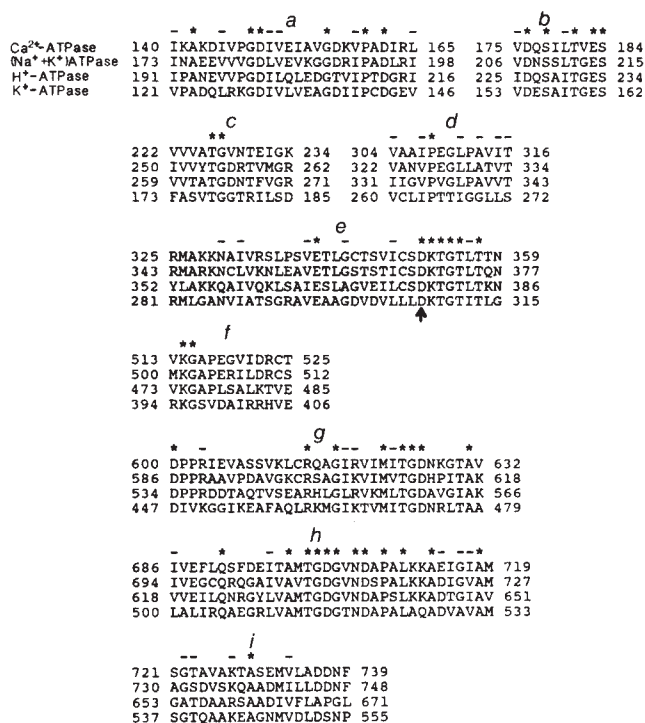


Fig. 3 Conserved sequences in four different ATPases having phosphorylated intermediates. The aspartyl residue that forms such an intermediate is marked by an arrow. Only sequences that are conserved in at least three of the enzymes are shown, and residues that are identical in all four enzymes are indicated by asterisks. Highly conserved replacements are indicated by a dash. Homology comparisons were made using the DIAGON program of Staden²⁸. The (Na⁺ + K⁺)ATPase sequence corresponds to the sheep kidney enzyme¹³, the Ca²⁺-ATPase sequence is from the cardiac muscle sarcoplasmic reticulum¹⁴, K⁺-ATPase refers to the product of the *KdpB* gene of *E. coli*¹², and the H⁺-ATPase refers to the yeast plasma membrane ATPase sequenced in the present work.

that are essential for their activity as cation pumps. Nine regions of homology were detected. Region *e* (Fig. 3) includes the aspartate that has been shown in animal and plant ATPases to be phosphorylated¹⁵. Nine amino acids surrounding this residue are identical in the yeast and animal ATPases, whereas the bacterial enzyme shares only seven identical amino acids. Region *f* contains the lysine residue which in the animal enzymes is labelled by fluorescein 5'-isothiocyanate and seems to form part of the ATP-binding site¹⁶. Four amino acids around this residue are identical in the yeast and animal ATPases, whereas only two are conserved in the bacterial enzyme. Region *h* has the greatest homology, with 12 identical amino acids conserved in each of the four ATPases. With the exception of regions *d* and *i*, homology is concentrated around polar residues, mainly aspartic acid (8 cases) and threonine (7 cases), with the small hydrophobic amino acids glycine and alanine often conserved in neighbouring positions. In general, the yeast ATPase is more related to the animal enzymes than to the bacterial one. These data support the notion¹²⁻¹⁶ that all the ATPases with a phosphorylated intermediate evolved from a common ancestor protein. It is tempting to speculate that this primitive ATPase operated as an H⁺ pump which later evolved the capability for Na⁺, K⁺ and Ca²⁺ transport¹⁷.

No sequence homology was detected between the yeast plasma membrane ATPase and the protein sequences of the H⁺-pumping ATPases of bacteria, mitochondria and chloroplasts (F₀-F₁ATPases; ref. 1). The lack of homology between these F₀-F₁ATPases and the yeast plasma membrane H⁺ pump

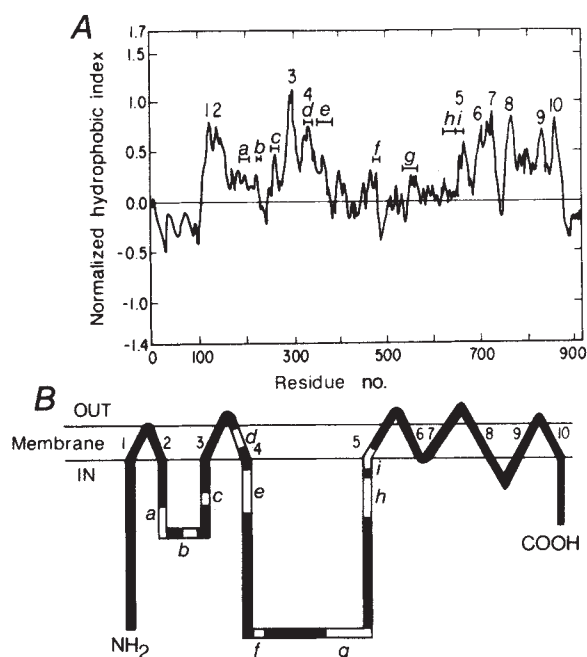


Fig. 4 **A**, Hydrophobicity profile of the yeast plasma membrane ATPase. Hydrophobicity values²⁹ were normalized to a mean value of 0.0 and standard deviation 1.0 and averaged over spans of 21 amino acids, as suggested by Eisenberg³⁰. Regions *a-i* correspond to the conserved sequences of Fig. 3; regions 1-10 are the hydrophobic stretches that are candidates for membrane-spanning sequences. Hydrophilic regions have negative values. **B**, Model of the transmembrane structure of the ATPase. The length of the bar representing the polypeptide chain is approximately proportional to the number of amino acids in every domain. The N-terminal domain (NH₂) contains 115 amino acids and is highly hydrophilic, without any of the features of a signal peptide. Two hydrophobic stretches of ~21 amino acids each, in the form of an α -helix, may allow the protein to cross the membrane twice, placing the N-terminal domain and the next hydrophilic domain of ~130 amino acids on the same side of the membrane. This hydrophilic region contains the first three conserved sequences of Fig. 3. Hydrophobic stretches 3 and 4 may represent two more membrane-spanning domains, and would place the large hydrophilic domain of ~310 amino acids on the same side of the membrane as the two hydrophilic domains that are more proximal to the amino-terminus. The hydrophilic central domain contains five of the conserved sequences of Fig. 3; as this segment also contains the phosphorylation site and the ATP-binding site, it should be exposed on the cytoplasmic side of the membrane. The conserved region *d* (corresponding to most of hydrophobic stretch 4) is close to the phosphorylation site, and has been proposed to constitute an energy-transducing pathway into the transmembrane domain that is involved in cation transport¹³. The last six hydrophobic stretches may also span the membrane. If so, the hydrophilic C-terminal domain (COOH) of ~46 amino acids would be on the cytoplasmic side, leaving very little of the enzyme exposed to the external medium.

suggests that the capability for ATP-driven proton transport arose in two independent ways. In fact, the two groups of ATPases are very different mechanistically, that is, the F₀-F₁ATPases do not form a phosphorylated intermediate. The F₀-F₁ATPases are ideally suited for oxidative and photosynthetic phosphorylation because the stoichiometry of 3 H⁺ per ATP¹⁸ allows an easy reversibility for ATP synthesis. On the other hand, fungal and plant plasma membrane ATPases have a stoichiometry of 1 H⁺ per ATP and are much more difficult to reverse, operating physiologically in the direction of ATP hydrolysis¹.

The hydrophobicity profiles of (Na⁺ + K⁺)ATPase¹³, Ca²⁺-ATPase¹⁴ and K⁺-ATPase¹² are very similar to that of the H⁺-

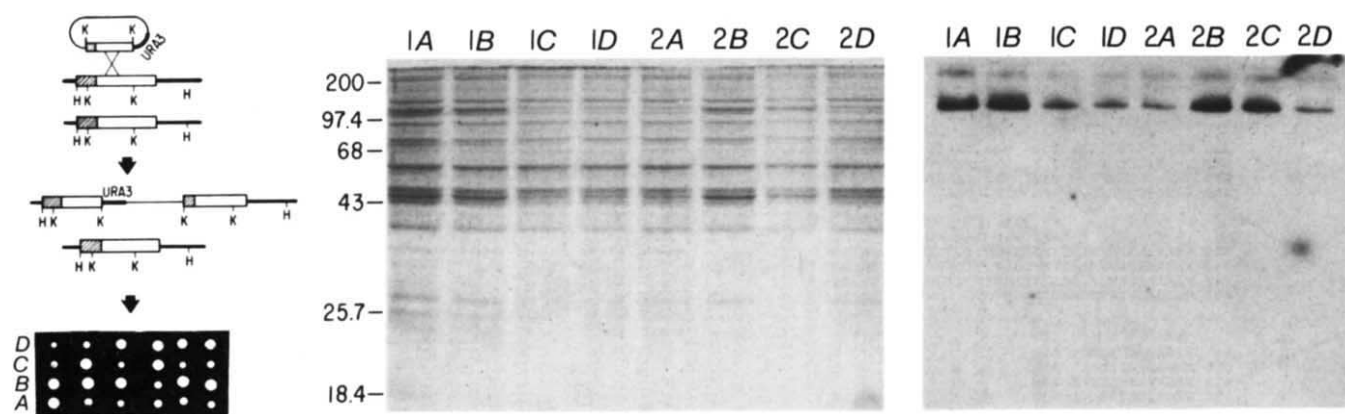


Fig. 5 Disruption of the ATPase gene. On the left of the figure is a diagram of the events leading to disruption of the gene. A Ura^- diploid strain

$$\left(\frac{a \text{ } ura3-52 \text{ } PMA1}{\alpha \text{ } ura3-52 \text{ } PMA1} \right)$$

was transformed with an integrating plasmid consisting of pUC18 (ref. 13), the yeast *URA3* gene and the *KpnI* fragment of the *PMA1* (see Fig. 1A). Integration at the ATPase locus was directed by cutting with *ClaI*. Thick and thin lines correspond to plasmid and yeast DNA, respectively. The open bar represents the ATPase coding region and the dashed bar the 5'-flanking region up to the *HindIII* site. H, *HindIII*; K, *KpnI*. Tetrads obtained by sporulation of this diploid are shown below the diagram. Electrophoretic analysis of the yeast colonies (A-D) from the first two tetrads (left to right) is shown on the right. Cells were grown in minimal medium with glucose, histidine and lysine. In the case of the large colonies (1A, 1B, 2B and 2C), uracil was also included. Cells were harvested at $A_{660} = 0.2-0.4$ and a total membrane fraction was prepared by homogenization with glass beads and differential centrifugation³¹. Assay of protein concentration (using the Bio-Rad Protein Assay Reagent) and plasma membrane ATPase activity³¹, and SDS-polyacrylamide gel electrophoresis were performed with this fraction. The gel on the left is stained for protein with Coomassie blue (12 μ g protein per lane), that on the right is a Western blot with antibody (purified by affinity chromatography) against the yeast plasma membrane ATPase (2 μ g protein per lane). The positions of the M_r markers are indicated. The growth rates and specific plasma membrane ATPase activities of the different colonies were as follows:

Colony	1A	1B	1C	1D	2A	2B	2C	2D
Growth rate (h^{-1})	0.32	0.32	0.11	0.11	0.12	0.32	0.30	0.11
Plasma membrane ATPase activity (units per mg protein)	0.19	0.17	0.07	0.06	0.07	0.16	0.15	0.07

ATPase (Fig. 4). Instead of the six hydrophobic stretches of the yeast enzyme (Fig. 4, regions 5-10), the animal ATPases seem to have only four and the bacterial ATPase just two. The lower relative molecular mass of the bacterial enzyme could be explained by a deletion of the carboxy-terminal segment. The hydrophobic stretches 1, 2, 5-7 and 9 of the yeast ATPase contain aspartic or glutamic acid residues which could participate in transmembrane proton transport¹⁹. There is a single potential N-linked glycosylation site (Asn-Trp-Thr) situated between hydrophobic stretches 9 and 10 and, in the working model of Fig. 4B, facing the external medium.

We propose the designation *PMA1* for the plasma membrane ATPase locus characterized here. The *PMA1* gene was localized to a subset of yeast chromosomes (VII and XV) by hybridizing the 5-kb *HindIII* fragment (see Fig. 1A) to yeast chromosomal DNA molecules separated by orthogonal-field-alternation electrophoresis^{20,21}. For exact localization we integrated the *URA3* gene at the *PMA1* locus using the *PMA1* homology to direct integration in strain R757 (*MAT α his4-15 lys9 ura3-52*). Mitotic mapping indicated that *PMA1* is located on the same arm of chromosome VII as *LEU1*. The gene was then mapped meiotically by crossing the *URA3-PMA1* transformant to S2021B (*MAT α his4 ura3 trp5 leu1 gal1 gal2*). All 32 resulting tetrads were of parental ditype for *URA3* and *LEU1*, showing that *PMA1* is linked to *leu1*, on chromosome VII. Comparison of the restriction map of the *PMA1* clone (Fig. 1A) with the map published for the *LEU1* region²² suggested an overlap between the *PMA1* and *LEU1* clones. We confirmed this overlap at the nucleotide level by sequencing downstream of the ATPase gene (Fig. 2) and found a region which has 425 nucleotides in common

with a sequence 5' to *LEU1*. Based on this information, the two genes are transcribed in the same direction, with the start codon of *LEU1* located 1.2 kb downstream of the stop codon of *PMA1* (Fig. 1A). Inspection of the genetic map in this region reveals a cluster of antibiotic-resistance genes closely linked to *LEU1*. If these resistance genes are in the *PMA1* locus, the situation would be analogous to that found in *E. coli* where several mutations affecting the resistance to antibiotics are located in the genes encoding the membrane ATPase²³.

The *EcoRI* and *KpnI* fragments shown in Fig. 1A were used for gene disruption experiments. Each fragment was cloned into a derivative of pUC18 (ref. 13) containing the *URA3* gene; the resulting plasmids are integrating vectors that can be directed to insert into the chromosome²⁴ at the *PMA1* locus by linearizing them with an enzyme that cuts uniquely in the yeast fragment on the plasmid (*HpaI* for the *EcoRI* fragment and *ClaI* for the *KpnI* fragment; see Fig. 1A). Each of these linear plasmids was used to transform a diploid strain homozygous for the *ura3-52* mutation to *Ura*⁺. Integration of the *KpnI* fragment results in a *PMA1* gene with an intact coding region and a deleted 5' noncoding region (Figs 1A, 5). The *KpnI* deletion begins 370 nucleotides upstream from the initiation codon and probably removes promoter sequences that are necessary for normal expression. After sporulation and dissection, the *KpnI*-transformed diploid gave rise to four viable spores. However, the two *Ura*⁺ spores (containing the *KpnI* deletion) always grew more slowly than the *Ura*⁻ sister spores (Fig. 5). The *Ura*⁺ cells had a growth rate in liquid medium ~35% of that of the *Ura*⁻ cells, and exhibited a corresponding reduction of the activity of the plasma membrane ATPase (Fig. 5). In each tetrad the

two Ura⁺ spores showed a 2:2 segregation for reduced immunological reactivity with the anti-ATPase antibody on Western blots and a reduced intensity of the M_r 100,000 (100K) band representing the ATPase on Coomassie blue-stained gels. The cosegregation of the reduced amounts of the ATPase with the genetic alteration in the *PMA1* gene supports the conclusion that the *PMA1* gene encodes the ATPase. The *EcoRI* subfragment is wholly internal to the *PMA1* gene, therefore its integration leads to disruption of the coding region (Fig. 1A). After sporulation and tetrad dissection, the *EcoRI*-transformed diploid gave only two viable spores in each tetrad and the surviving spores were all Ura⁻. These results show that cells lacking a functional *PMA1* gene are not viable.

The requirement of a functional plasma membrane ATPase for viability could be explained in several ways. One likely possibility is that this H⁺ pump is required for active nutrient transport and for pH regulation¹. In addition, the ATPase protein may be required to maintain integrity of the plasma membrane. Such possibilities can now be investigated by further mutational analysis of the cloned gene.

We thank Richard A. Young for the genomic yeast DNA library in λ gt11 and the bacterial strains Y1090 and Y1089; David Botstein for the YEp24 library; Mark Rose for the YCp50 library; and Rick Gaber, George Natsoulis and Gary E. Schull for technical advice. We also thank Gary E. Shull and David H. MacLennan for making available the sequences of the animal ATPases before publication. M.C.K.-B. was on leave from the Carlsberg Laboratory, Copenhagen, Denmark and R.S. (on leave from the Instituto de Enzimologia CSIC, Madrid, Spain) was supported by an EMBO fellowship. Antibody preparation was supported by a grant of the CAICYT (Spain) to R.S. This

work was supported by NIH grants CM 35010 and CA 34429 to G.R.F., who is an American Cancer Society Research Professor of Genetics.

Received 10 September; accepted 8 November 1985.

- Serrano, R. *Curr. Topics cell. Regul.* **23**, 87-126 (1984).
- Young, R. A. & Davis, R. W. in *Genetic Engineering: Principles and Methods* Vol. 7 (eds Setlow, J. K. & Hollaender, A.) 29-41 (Plenum, New York, 1985).
- Melpartida, F. & Serrano, R. *Eur. J. Biochem.* **116**, 413-417 (1981).
- Mihara, K. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4160-4164 (1980).
- Drescher, D. G. & Lee, K. S. *Analyt. Biochem.* **84**, 559-569 (1978).
- Thomas, R. E., Korzeniewski, D., Ryan, D. & Levin, W. *Archs Biochem. Biophys.* **192**, 524-532 (1979).
- Burnette, W. N. *Analyt. Biochem.* **112**, 195-203 (1981).
- Rott, R. & Nelson, N. *J. biol. Chem.* **256**, 9224-9228 (1981).
- Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning Techniques: A Practical Approach* (ed. Glover, D.) (IRL, Oxford, in the press).
- Norrander, J., Kempe, T. & Messing, J. *Gene* **26**, 101-106 (1983).
- Pikielny, G. W., Teem, J. L. & Rosbash, M. *Cell* **34**, 395-403 (1983).
- Hesse, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4746-4750 (1984).
- Shull, G. E., Schwartz, A. & Lingrel, J. B. *Nature* **316**, 691-695 (1985).
- MacLennan, D. H., Brandl, C., Korczak, B. & Green, N. M. *Nature* **316**, 696-700 (1985).
- Walderhaug, M. O., Post, R. L., Saccomani, G., Leonard, R. T. & Briskin, D. P. *J. biol. Chem.* **260**, 3852-3859 (1985).
- Farley, R. A. & Faller, L. D. *J. biol. Chem.* **260**, 3899-3901 (1985).
- Dame, J. B. & Scarborough, G. A. *J. biol. Chem.* **256**, 10727-10730 (1981).
- Berry, E. A. & Hinkle, P. C. *J. biol. Chem.* **258**, 1474-1486 (1983).
- Ovchinnikov, Y. A., Abdulaev, N. G. & Modyanov, N. N. *A. Rev. Biophys. Bioengng* **11**, 445-463 (1982).
- Schwartz, D. C. & Cantor, C. R. *Cell* **37**, 67-75 (1984).
- Carle, G. F. & Olson, M. V. *Nucleic Acids Res.* **12**, 5647-5664 (1984).
- Hsu, Y. P. & Schimmel, P. J. *J. biol. Chem.* **259**, 3714-3719 (1984).
- Downie, J. A., Gibson, F. & Cox, G. B. A. *Rev. Biochem.* **48**, 103-131 (1979).
- Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6354-6358 (1981).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Bencini, D. A., O'Donovan, G. A. & Wild, J. R. *Biotechniques* **2**, 1-70 (1984).
- Heffron, F., So, M. & McCarthy, B. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6012-6016 (1978).
- Staden, R. *Nucleic Acids Res.* **10**, 2951-2961 (1982).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
- Eisenberg, D. A. *Rev. Biochem.* **53**, 595-623 (1984).
- Serrano, R. *FEBS Lett.* **156**, 11-14 (1983).

Location of exit channel for nascent protein in 80S ribosome

R. A. Milligan & P. N. T. Unwin

Department of Cell Biology, Stanford University School of Medicine, Stanford, California 94305, USA

Ribosomes crystallize on endoplasmic reticulum membranes in oocytes of the southern Italian lizard, *Lacerta sicula*, during winter¹. Electron crystallographic studies of the crystals have been made to elucidate the arrangement of the ribosomal subunits on the membrane surface^{2,3}. We have now obtained more extensive and better ordered crystals of the same habit, grown from chick embryo ribosomes⁴, and report here on their native structure preserved by rapid freezing of the crystals in thin aqueous films. The three-dimensional map reveals new details of the protein and ribosomal RNA distribution within the ribosome. Most striking is a region of low density within the large subunit which extends from the subunit interface towards an area on the membrane-facing surface identified by others^{5,6} as the exit site of the nascent protein. This region of low density appears to delineate the path taken by the growing polypeptide through the ribosome to the external surface.

The chick embryo crystals (Fig. 1) are composed of ribosome tetramers in two opposite-facing layers, arranged in the two-sided plane group p4₂2 ($a = 593$ Å). Each of the layers, when viewed from the mid-plane of the crystal, presents a right-handed p4 lattice, which identifies their outermost surfaces as being equivalent to the surface which, in the lizard, attaches to the membrane³. Specimens were prepared for electron microscopy by rapid freezing in thin aqueous films^{7,8}, and images were recorded of the specimens tilted at angles of up to 50° with respect to the incident electron beam. The images were analysed by standard methods⁹ to provide a three-dimensional dataset of

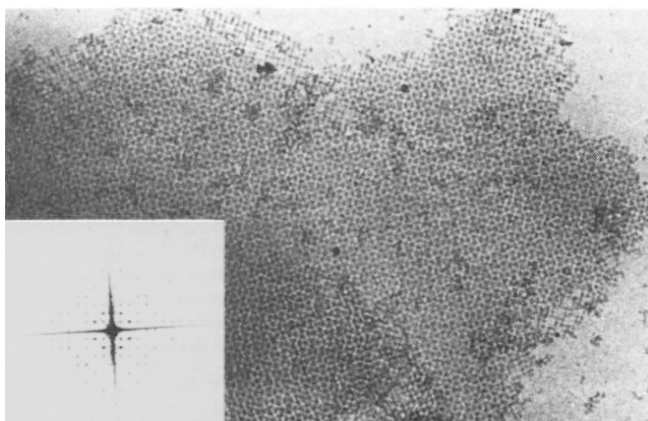


Fig. 1 Image of a ribosome crystal ($\times 37,500$) embedded in a thin film of frozen buffer, and optical diffraction pattern (inset) from an area containing $\sim 15 \times 15$ unit cells. The crystal was tilted at an angle of 10° to the incident electron beam.

Methods. Crystals, prepared as described in ref. 4, were washed free of mother liquor by low-speed centrifugation and resuspended in a solution in which ribosomes are potentially active (50 mM KCl, 10 mM MgCl₂, 0.5 mM EDTA, 0.5 mM dithiothreitol, 10 mM HEPES pH 7.2). They were applied in 5- μ l aliquots to amylamine-treated carbon-coated grids and, after the excess liquid had been blotted off, were quick-frozen by plunging the grids into liquid ethane slush^{7,8}. The grids were mounted under liquid nitrogen in a Philips cold holder, PW 6591/00, and examined at -160 °C using an EM400 electron microscope operating at 100 kV. Protection of the grid by an additional anti-contamination device²¹ was necessary during the 10 min after insertion. Only one image was recorded from each frozen crystal. Each area photographed received a total dose of 2-5 electrons per Å². Images were recorded at an electron optical magnification of $\times 19,500$ or $\times 15,200$ with an underfocus (2.0-2.5 μ m) such that the first zero of the contrast transfer function lay just outside the resolution ($1/55$ Å⁻¹) to which the diffraction patterns extended.

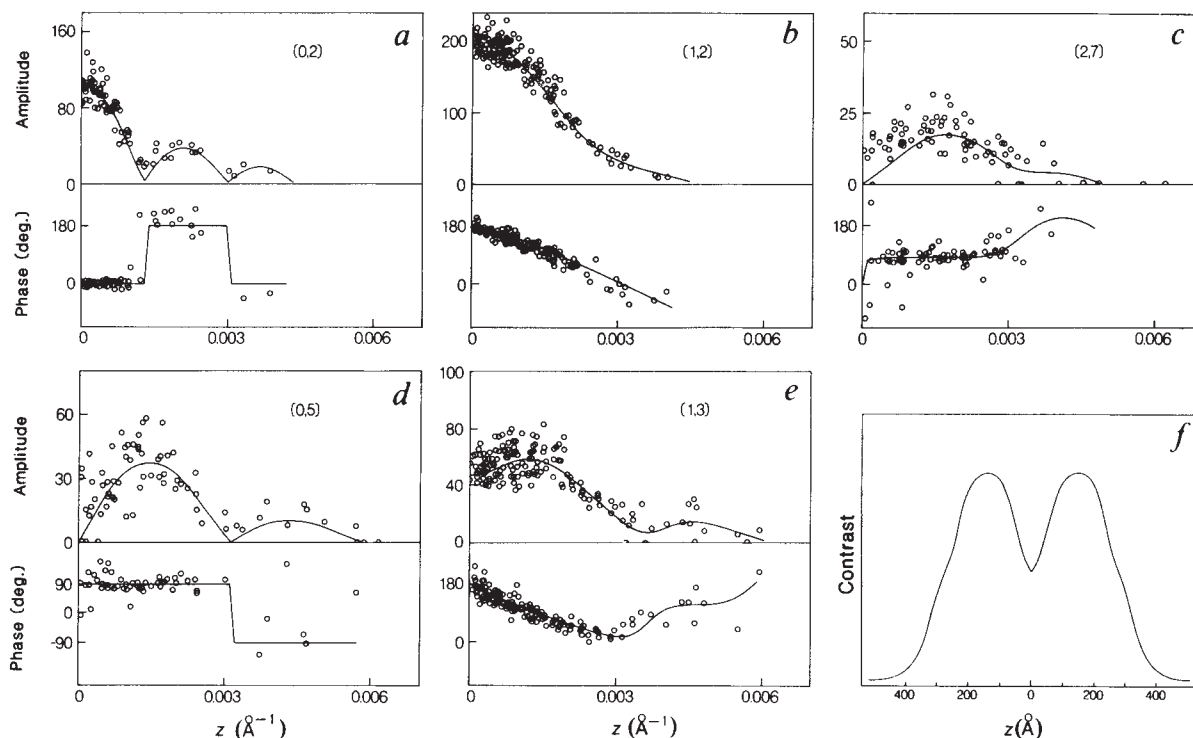


Fig. 2 *a-e*, Variations of amplitude and phase along 5 of the 42 lattice lines from which the three-dimensional map was synthesized (p42₁2 symmetry imposed). Open circles represent the experimental values obtained by computer processing of the images, solid lines the fitted curves. z^* is the distance from the mid-point of the lattice line. *f*, Plots of the contrast (maximum density minus minimum density in each section) through the map in the direction perpendicular to the plane of the crystal. The crests of the peaks which represent the two ribosome layers are separated by ~ 280 Å.

Methods. Computer processing of images was done in the standard way⁹. Highly ordered areas in the images were selected by optical diffraction and scanned with a Perkin-Elmer microdensitometer having step and sampling sizes of 25 μm , corresponding to 13 Å at the specimen. We calculated Fourier transforms (512×512) of the optical density arrays thus generated, and reciprocal lattices were fitted to the strong peaks. Amplitudes and phases of above-background peaks were extracted at the reciprocal lattice points to the resolution observed in the optical diffraction patterns. Several images of untilted crystals giving average phase residuals of $<25^\circ$, based on a comparison of individual 4-fold related peaks, were used for the initial data in the three-dimensional analysis. Thirty-nine images from tilted crystals (tilt angle range 4–50°) were used for the remaining data, the tilt axis and angle parameters being calculated from the distortion of the reciprocal lattice. The data were first combined assuming p4 symmetry, using a comparison range in z^* of 0.001 Å^{-1} . After we had ascertained that the ribosomes in the two crystal layers were equally well preserved—that is, showing the same essential features independent of the layer in question—the data were combined with p42₁2 symmetry. The average phase error for each image, based on a cumulative comparison of individual symmetry-related phases, was 22° . Smooth curves were fitted to the experimental values along each lattice line²² and the curves were sampled at regular intervals ($1/1,500 \text{ Å}^{-1}$) to provide the terms for Fourier synthesis. Terms along the 0,0 lattice line were not included. No corrections were made to compensate for the effects of the phase-contrast transfer function as the background amplitude in the computed transforms showed only a 10–20% change in the range of spatial frequencies over which data were collected. Sections through the map were traced onto plastic sheets which were stacked up to show the overall density distribution. The molecular boundary was chosen to be that at which ribosomes in a tetramer just make contact.

42 crystallographically independent lattice lines, extending to a resolution of ~ 55 Å (Fig. 2*a-e*). The map was calculated by Fourier synthesis of terms obtained by sampling at regular intervals the continuous variations of amplitude and phase along each of the lattice lines. A plot of the contrast through the structure normal to the plane of the crystal (Fig. 2*f*) shows the separation between the centres of mass of the two layers to be ~ 280 Å, in good agreement with the 266-Å spacing for this distance measured from low-angle X-ray diffraction patterns⁴.

Unlike structures determined from negatively stained specimens, where only stain-accessible surfaces are revealed, the three-dimensional map presented here (Fig. 3) gives a description of the native densities within the ribosome as the crystals have been preserved unstained and frozen in an appropriate ionic environment. Viewed with the membrane-facing surface at the bottom (Fig. 3*a, b*), the ribosome forms a rounded outline with two indentations (arrowheads in Fig. 3*a*) suggestive of a partitioning into a larger portion (the large subunit) closest to the 4-fold axis of the tetramer, and a smaller ~ 210 -Å-long portion (the small subunit) at the periphery. Viewed with the membrane-facing surface uppermost (Fig. 3*c*), the ribosome

surface comes closest to the observer in a region of the large subunit (M in Fig. 3*c*) identified previously as the salt-sensitive membrane attachment site². Within the ribosome there is a dense core of irregular shape (shading in Fig. 3*b, c*), which must be composed largely of rRNA (the most strongly scattering component), because essentially only this density is visualized when contrast from the protein is matched out in a solution containing 10% metrizamide (data not shown). Apparently, much of the interface area between the two subunits is composed of rRNA.

These details are in good agreement with those found in earlier studies. The assignment for the ribosomal subunits corresponds to that determined by comparing maps from arrays of tetramers in negative stain with and without the small subunit present³. As before, it appears that the subunits lie approximately side-by-side on the endoplasmic reticulum membrane and are attached to it through a part protruding from the large subunit. Consistent with contrast-matching studies at lower resolution using neutron^{10,11} and electron¹² irradiation, the rRNA and protein are distributed inhomogeneously, with rRNA-rich regions concentrated in the interior of the particle.

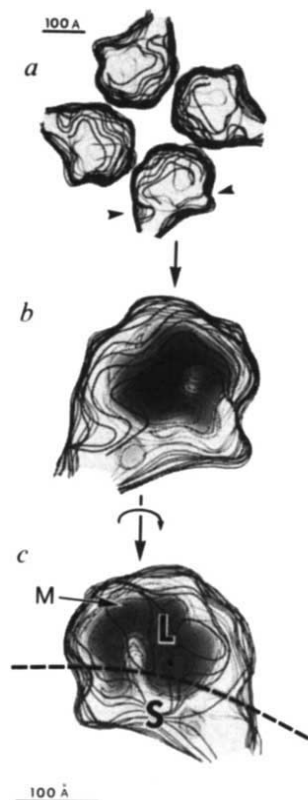


Fig. 3 Three-dimensional contour map showing: *a*, a ribosome tetramer viewed from the mid-plane of the crystal so that the cytoplasmic-facing surface of the ribosome (see text) is closest to the observer; *b*, a single ribosome viewed from the same direction; *c*, a single ribosome viewed so that the membrane-facing surface is closest to the observer. In *a* the molecular boundary alone is shown; contours are incomplete at the periphery, where tetramers in a single layer of the crystal make contact, and at the top (clearer contours) where the two layers of ribosomes in the crystal touch. Arrowheads in *a* and the broken line in *c* follow the surface indentations which appear to identify the boundary between the large (L) and small (S) subunits. The shading in *b* and *c* indicates the distribution of the high-density (rRNA-rich) regions; it makes up 25% of the total volume, or ~65% of the rRNA if one were to assume complete partitioning of the rRNA and protein. M is the part of the large subunit closest to the membrane (see also Fig. 4). Negative stain has been shown to penetrate parts of the structure adjacent to the central region of low density in *c*^{2,23}.

The feature of particular interest, revealed by this higher resolution analysis, is a narrow elongated region of low density passing through the RNA-rich core at an angle of ~30° to the presumed subunit interface (Fig. 4). This feature does not span the ribosome completely, but appears to originate internally, near the subunit interface, and terminate at a point on the surface of the large subunit close to the membrane attachment site—a total distance of 150–200 Å. Based on correlations made between ribosomes analysed in crystals and isolated ribosomes studied by immunoelectron microscopy⁶, the point of intersection with the ribosome surface is close to the exit site of the nascent protein. The low-density region therefore appears to make a direct line between the site at which the growing polypeptide originates, the subunit interface (see refs 13, 14), and the place where the polypeptide emerges from the ribosome. We conclude that the low-density region represents the location of a channel or deep groove in the large subunit along which the growing polypeptide is able to pass.

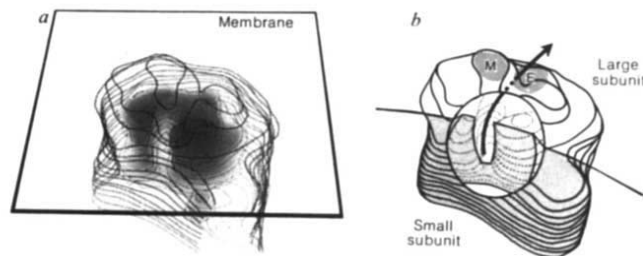


Fig. 4 *a*, Oblique view of the ribosome with the membrane-facing surface towards the observer; *b*, schematic representation of the details revealed. The arrow in *b* indicates the presumed pathway of the polypeptide chain, based on the distribution of low densities in *a*. The point of emergence from the large ribosomal subunit appears to be close to both the membrane attachment site (M)² and the exit site (E) determined by immunoelectron microscopy^{5,6}.

Our results support the deductions made about the location of the channel by Lake and colleagues⁶, based on immunolabelling of nascent protein attached to isolated eukaryotic ribosomes. The channel is clearly long enough to account for the observation that up to 39 amino-acid residues are protected from proteolytic enzymes by the large subunit^{15–17}. Given that the channel is at least 150 Å long, the enclosed polypeptide is presumably in an extended conformation, implying a bore size of not greater than ~10 Å. A larger diameter would allow formation of secondary structure, so that a much greater portion of the nascent protein than has been found experimentally would be protected. As details of dimensions as small as 10 Å are beyond the resolution of the map, the most likely explanation for the prominence of the low-density path is that the channel is surrounded by ribosomal protein and that the two together make up the observed extensive region of low electron-scattering density. The orientation of the channel seems an ideal one to deliver the polypeptide from the site of translation between the subunits to a point on the external surface adjacent to the membrane where it is optimally located to interact with components of the translocation apparatus such as the signal recognition particle¹⁸, docking protein¹⁹ and the ribophorins²⁰.

We thank Jacques Dubochet, David Grano and Jean Lepault for their help. This research was made possible by grants and fellowships from the NIH, SERC/NATO, EMBO and the Royal Society.

Received 11 September; accepted 13 December 1985.

1. Taddei, C. *Expl Cell Res.* **70**, 285–292 (1972).
2. Unwin, P. N. T. *Nature* **269**, 118–122 (1977).
3. Unwin, P. N. T. *J. molec. Biol.* **132**, 69–84 (1979).
4. Milligan, R. A. & Unwin, P. N. T. *J. Cell Biol.* **95**, 648–653 (1982).
5. Bernabeu, C. & Lake, J. A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3111–3115 (1982).
6. Bernabeu, C., Tobin, E. M., Fowler, A., Zabin, I. & Lake, J. A. *J. Cell Biol.* **96**, 1471–1474 (1983).
7. Lepault, J., Booy, F. P. & Dubochet, J. *J. Microsc.* **129**, 89–102 (1983).
8. Milligan, R. A., Brisson, A. & Unwin, P. N. T. *Ultramicroscopy* **13**, 1–10 (1984).
9. Amos, L. A., Henderson, R. & Unwin, P. N. T. *Prog. Biophys. molec. Biol.* **39**, 183–231 (1982).
10. Serdyuk, I. N., Grenader, A. K. & Zaccari, G. *J. molec. Biol.* **135**, 691–707 (1979).
11. Stuhmann, H. B. *et al. J. molec. Biol.* **119**, 203–212 (1978).
12. Kühlbrandt, W. & Unwin, P. N. T. *J. molec. Biol.* **156**, 431–448 (1982).
13. Liljas, A. *Prog. Biophys. molec. Biol.* **40**, 161–228 (1982).
14. Noller, H. F. A. *Rev. Biochem.* **53**, 119–162 (1984).
15. Malkin, L. I. & Rich, A. *J. molec. Biol.* **26**, 329–346 (1967).
16. Blobel, G. & Sabatini, D. D. *J. Cell Biol.* **45**, 130–145 (1970).
17. Smith, W. P., Tai, P.-C. & Davis, B. D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5922–5925 (1978).
18. Walter, P. & Blobel, G. *Nature* **299**, 691–698 (1982).
19. Meyer, D. I., Krause, E. & Dobberstein, B. *Nature* **297**, 647–650 (1982).
20. Kreibich, G., Ulrich, B. L. & Sabatini, D. D. *J. Cell Biol.* **77**, 464–487 (1978).
21. Homo, J.-C., Booy, F., Labouesse, P., Lepault, J. & Dubochet, J. *J. Microsc.* **136**, 337–340 (1984).
22. Agard, D. A. *J. molec. Biol.* **167**, 849–852 (1983).
23. Milligan, R. A. thesis, Stanford Univ. (1985).

The Precambrian–Cambrian boundary and geochemical perturbations

Hsü *et al.*¹ present intriguing data on geochemical perturbations in two Lower Cambrian sections in China. At or near the base of the lowest trilobite-bearing formations in Yunnan and the Yangtze Gorges, they found a clay layer containing 5% more negative $\delta^{13}\text{C}$ and increased iridium, osmium, and gold contents relative to strata above and below the clay. A more negative $\delta^{18}\text{O}$ perturbation was found a few centimetres upsection. Figures 1, 2 and 3 in ref. 1 show that the $\delta^{13}\text{C}$, Ir, Os, and Au anomalies occur right at the China C marker while the text indicates that the anomalies occur above the marker. Hsü *et al.*, using the China C marker as the Precambrian–Cambrian boundary (apparently using the palaeontological criterion of placing the boundary at the base of the formation with the lowestmost trilobite), concluded that the $\delta^{13}\text{C}$ perturbation reflected a fertility crisis (due to a phytoplankton mass mortality or mass extinction brought about by a catastrophic event, hence the iridium anomaly) immediately preceding the Cambrian explosion of invertebrate evolution.

Regardless of the precise position of the geochemical changes with respect to the China C marker, these data do not coincide with any measurable mass extinction or phytoplankton mass mortality, and do not precede the Cambrian explosion. They occur during the explosion, when Metazoa are undergoing the most magnificent radiation in their history.

The China C marker in Yunnan and the Yangtze Gorges, demarcates a lithologic boundary and the base of the lowest formation containing trilobites¹. An unconformity has been reported at this boundary² although recent interpretations favour a facies change reflecting a transgression³. Below China C and the horizons containing the geochemical anomalies, abundant and diverse pre-trilobite faunas are well known. In Yunnan at the Kunyang Phosphate Mine in strata older than the China C marker, some 74 invertebrate species have been noted³. There is general consensus among researchers on the Precambrian–Cambrian boundary, that it be placed at the lowest known appearance of diverse shelly fossils with a good potential for correlation⁴. In China, at Yunnan, this corresponds to the base of unit 7 of the Zhongyicun member⁴ (at China B). It must be noted that even below this datum, shelled invertebrates, albeit of comparatively low diversity, are found³. Similar faunal sequences and relationships are known from the Soviet Union, Mongolia, Mexico, Newfoundland, Spain, England, and elsewhere⁵. These

pre-trilobite biotas, composed of millimetre-sized calcium carbonate and phosphatic shells (small shelly fossils), are so distinctive and widespread that the term Tommotian Stage is being used to encompass the strata containing them⁶. Phytoplankton represented by acritarchs, do not appear to suffer any mass mortality or extinction at or near the China C horizon; indeed, they underwent a radiation that parallels the animals⁷.

It is probable that the geochemical perturbations have nothing whatsoever to do with the first trilobites as suggested by Hsü *et al.* Trilobites, probably older than those known from Yunnan³ are found in Morocco, with *Hupetina* occurring below *Lamdelella* and *Fallotaspsis*⁸. Unfortunately, the interregional correlation of earliest Cambrian trilobites is imprecise and does not agree with schemes based on archaeocyathids. Trace fossils like *Rusophycus* and *Cruziana*, attributable to trilobites but more conservatively interpreted as having been produced by arthropods with sclerotized legs that behaved like trilobites, are known from strata older than China C^{3,9}. It is unnecessary to invoke the geochemical perturbations as an external, environmental stimulus for the early evolution of shelled animals. Evolutionary theory is consistent with the occurrence of a broad diversification in a world of unfilled or partially filled niches¹⁰.

An iridium spike has also been reported from the USSR's stratotype for the Precambrian–Cambrian boundary. The base of bed 8 (Pestrotsvet Formation) at Ulakhan–Sulugur, Siberia, contains a 6-fold iridium enrichment, but no anomalous trace element shifts (Os and Au not reported)¹¹. No $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ data were noted. The base of bed 8, which rests with slight erosional unconformity on bed 7, is roughly correlated on palaeontological grounds with unit 7 of the Zhongyicun¹¹ or China B. However, iridium analyses of the same bed 8 samples by Orth *et al.*¹² (and C. J. Orth, personal communication) failed to detect iridium enrichment.

Some entirely different conclusions can be drawn from Hsü *et al.* (1) iridium spikes are more common in the geological record than previously thought and some, possibly most, are not associated with mass extinctions; (2) other geochemical perturbations ($\delta^{13}\text{C}$, $\delta^{18}\text{O}$, Os, Au) hitherto associated with iridium and catastrophes also occur without catastrophic effects on biota; (3) $\delta^{13}\text{C}$ excursions appear to have been common in the Vendian and early Cambrian (A. H. Knoll, personal communication). These might represent the response of the oceans to changes brought about by widespread biomineralization and the rapid diversification and proliferation of phytoplankters; and (4) geochemical perturbations across the Pre-

cambrian–Cambrian transition may be associated with small-scale stratigraphic breaks and could be used as a tool in recognizing such hiatuses. As suggested by Hsü *et al.*, these geochemical perturbations may be useful as chemochronostratigraphic signals for the correlation of widely separated regions; but they may not necessarily be coincidental with biostratigraphic boundaries.

This work was supported by NSF grants EAR82-05809 and 85-08984.

STANLEY M. AWRAMIK

Department of Geological Sciences
Preston Cloud Research Laboratory,
University of California,
Santa Barbara,
California 93106, USA

1. Hsü, K. J., Oberhänsli, H., Gao, J. Y., Sun Shu, Chen Haihong & Krähenbühl, U. *Nature* **316**, 809 (1985).
2. Liu Hung-Yun, Sha Ching-An & Hu Shih-Ling *Sci. Sin.* **16**, 266 (1973).
3. Luo Huilin *et al.* *Sinian–Cambrian Boundary Stratotype Section at Meishucun, Jinning, Yunnan, China* (People's Publishing House, Yunnan, 1984).
4. Cowie, J. W. *Episodes* **8**, 93 (1985).
5. Rozanov, A. Yu. *et al.* *Lower Cambrian Stage Subdivision, Stratigraphy* (Acad. Sci. USSR, Nauka, 1984).
6. Bengtson, S. *Lethaia* **14**, 372 (1981).
7. Vidal, G. & Knoll, A. H. *Nature* **297**, 57–60 (1982).
8. Szduy, K. *Geol. Mag.* **115**, 83 (1978).
9. Glaessner, M. F. *The Dawn of Animal Life* (Cambridge University Press, New York, 1984).
10. Valentine, J. W. in *Phanerozoic Life, Patterns and Process* (eds Raup, D. M. & Jablonski, D.) (Springer, Berlin, in the press).
11. Nazarov, M. A., Barsukova, L. D., Kolesov, B. M. & Alekseev, A. S. *Lunar planet. Sci.* **14**, 546 (1983).
12. Orth, C. J., Gilmore, J. S. & Knight, J. D. *Lunar planet. Sci.* **16**, 631 (1985).

THE paper by Hsü *et al.*¹ reporting geochemical anomalies, specifically with regard to $\delta^{13}\text{C}$ and an iridium spike, near the base of the Cambrian in China commands attention. There have already been hints of major negative shifts in carbon isotope values near the Precambrian–Cambrian transition²; the data in ref. 1 indeed may point to a dramatically changing oceanic productivity. Extraterrestrial agencies are often invoked as an explanation of iridium anomalies³, an option apparently followed by Hsü *et al.*, but such an enrichment pattern in a black shale which appears to have been deposited in anoxic or dysaerobic conditions⁴ may also be of mundane origin.

The principal point we wish to take issue with, however, is the vigorously argued connection between these geochemical anomalies and the “Cambrian explosion”. There are many triggering mechanisms that have been invoked to explain this dramatic evolutionary radiation of metazoans near the beginning of the Cambrian (the “cropping hypothesis”⁵ referred to by Hsü *et al.* being only one of them), but Hsü *et al.*'s proposal is the only one we know of that would place the triggering event well after the “explosion”. The current major disagreements with regard to the global correlation of the Chinese Precambrian–Cambrian boundary sections

may be responsible for clouding the issue, but the precise age of the Chinese sections is not material to this argument. Whether the phosphatic limestones and dolomites underlying the black shales in China are older than the Siberian Tommotian Stage (as some say) or younger (as others maintain), they are in fact richly fossiliferous⁴, littered with shards from the "Cambrian explosion". According to Hsü *et al.*, the fact that the first trilobites in the area are found close to the anomaly cannot be coincidental. Why this emphasis on trilobites? Trilobites are large and an obvious component of many Cambrian shelly assemblages. Thus they earn their position as the most well-known Cambrian fossils, but they probably played a fairly subordinate role in most Cambrian ecosystems. Palaeontologists have long realized that there is much more to the "Cambrian explosion" than the appearance of trilobites, and that in any event the trilobites were by no means the first entrants.

Anomalies of the kind investigated by Hsü *et al.* will probably be of vital importance in regional correlations. They may reflect substantial crises in marine ecosystems. What is inadmissible is that these particular Chinese anomalies should be taken as the signature of a mass extinction that provided the ecological vacuum for the "Cambrian explosion". One might as well invoke the assassinations in Sarajevo as the likely cause of the Trojan Wars.

SIMON CONWAY MORRIS

Department of Earth Sciences,
University of Cambridge,
Cambridge CB2 3EQ, UK

STEFAN BENGTON

Institute of Palaeontology,
Box 558,
S-751 22 Uppsala, Sweden

1. Hsü, K. J. *et al.* *Nature* **316**, 809–811 (1985).
2. Veizer, J., Holser, W. T. & Wilgus, C. K. *GEOCHIM. COSMOCHEM. ACTA* **44**, 579–587 (1980).
3. Smit, J. & Romein, A. J. T. *Earth planet. Sci. Lett.* **74**, 155–170 (1985).
4. Luo, H. *et al.* *Sinian–Cambrian Boundary Stratotype Section at Meishucun, Jinning, Yunnan, China* (People's Publishing House, Yunnan, 1984).
5. Stanley, S. M. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1486–1489 (1973).

HSÜ REPLIES—I am pleased to see that Conway Morris and Bengtson accepted our discovery of a carbon-isotope anomaly at the China-C marker, the base of the first trilobite-bearing formation. They also agreed that our data may point to dramatically changing ocean productivity. They assert, however, that the Cambrian explosion occurred before, not after, the Strangelove ocean. They present little evidence to support their assertion. Awramik, who has also visited the Chinese sections, does present his arguments, which are well known to us.

The crux of the matter is the definition of the Cambrian explosion. Palaeontologists,

especially those on the Precambrian–Cambrian Boundary Working Group of IUGS, have been arguing for years the relative significance of several evolutionary events during the late Precambrian and early Cambrian. They have not yet reached consensus on whether the China A (first appearance of small shelly fauna), B (diversification of the same), or C (first appearance of trilobites) signifies the Cambrian explosion. Conway Morris and Bengtson, as well as Awramik, seem to have opted for China B. We have found no geochemical anomalies at either the China A or China B datum horizons, but a remarkable set of anomalies at the China C datum. This is the reason why we found the opinions of our Chinese friends more acceptable. Zhang Quingwin of Academy of Geological Sciences, Beijing, has suggested China C as the major biostratigraphical marker, and he presented his arguments at the Conference on Ocean Crises in Earth History, held at Gwatt, Switzerland this year. When I mentioned to J. W. Cowie, Chairman of the IUGS Commission on Stratigraphy and former chairman of the boundary group, the suggestion that the base of Cambrian be placed at the first appearance of trilobite datum, he did not regard the idea outrageous. Accepting this definition, the assassination of Sarajevo was the cause of the Great War, not of the Trojan War.

KENNETH J. HSÜ

Geological Institute
ETH Zurich, Switzerland

Highly conducting acetylene–CO copolymers and limitations of the soliton model

RECENTLY, Chien *et al.*^{1,2} described a proposed copolymer composition of carbon monoxide and acetylene on which they based the following generalized conclusions: (1) there is no need to invoke the soliton hypothesis to explain a wealth of results of physical measurements on polyacetylene; and (2) only short conjugation lengths are required for high conductivity on doping.

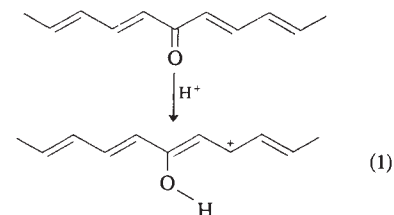
Here we point out that the conclusions of Chien *et al.*^{1,2} are not consistent with their experimental facts, and we present an alternative explanation.

Most fundamentally disturbing about the hypothesis of Chien *et al.* is the lack of a conclusive demonstration that the CO has been distributed homogeneously throughout the polyacetylene. Without convincing evidence that the polymer consists of short segments of polyenes regularly interrupted by carbonyls, the 'copolymer' in this study is not a valid model on which to base a new theory of conduction. To assume regular and discrete lengths for the polyene segments based on the ratio

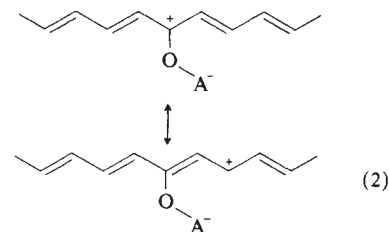
of CO to acetylene is naive, and unfortunately the data of Chien *et al.* do not sufficiently document the homogeneous distribution of CO in their polyacetylene. We present here our attempts to identify the product of Chien and Babu's reaction based on the information available from refs 1–3.

Their basic premise that the CO groups prevent the unpaired spins (solitons) from moving is not substantiated by their data. An interpretation of their data is facilitated by differentiating between inter- and intramolecular mobility. The narrow electron spin resonance (ESR) linewidth (1.1 G) and the T_1 and T_2 values indicate that the spins are at least as mobile on an intramolecular level as in polyacetylene, implying that the copolymers indeed do contain a significant extent of conjugation. Moreover, as the conductivity of the doped 'copolymer' is relatively high, one can infer that the inclusion of CO groups does not inhibit inter-chain transport.

How might extended conjugation in a polymer purported to contain frequently interrupted polyene segments come about? In considering plausible structures for a 'copolymer' of CO and acetylene, it can be seen that insertion of a CO-derived moiety need not interrupt conjugation:



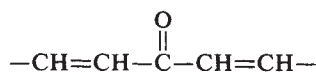
The likelihood of keto-enol tautomerization (enhanced by the expected high basicity for a polyene–CO structure) is supported by Chien and Babu's elemental analyses which show an excess of hydrogen. The isomerization from a ketone to an enol represents a mechanism by which CO inclusion in a polyene would not necessarily shorten the conjugation length. The existence of C–OH units could be verified by detailed infrared studies in the spectral region of the OH stretch (3,300–3,500 cm⁻¹). An equally plausible structure considers the polymer in its doped, ionized state wherein the carbonyl (if present) would act as a source of mobile positive solitons:



where A is Lewis acid-derived, for example, from AsF₅.

Chien and Babu³ have attempted to characterize the material by magic-angle

spinning ^{13}C -NMR and by infrared spectroscopy. It is impossible to extract quantitative information about CO inclusion from their NMR data because the weak signal at 178 p.p.m. (parts per 10^6) which they attribute to carbonyl groups, is actually assigned to the sample holder (P. Bernier, personal communication; see also ref. 4). Moreover, comparison with analogous unsaturated carbonyl-containing compounds indicates that the carbonyl resonance of such a proposed CO-acetylene copolymer should display a chemical shift at ~ 188 p.p.m. (there are no signals in this region in the spectra obtained from the 'copolymers'). (Note, for example, the ^{13}C -carbonyl chemical shifts of 188.2 p.p.m. for tropone (2,4,6-cycloheptatrienone)⁵ and 188.4 p.p.m. for 1,5-diphenyl-1,4-pentadien-3-one; Sadler spectrum no. 1058C.) The weak infrared lines at 1,670 and 1,720 cm^{-1} were attributed by Chien *et al.* to



and



respectively. Given the large extinction coefficient expected for a $\text{C}=\text{O}$ bond stretch, their signal strength is too weak to account for the proposed 13% $\text{C}=\text{O}$ in the sample³. (The structure given in ref. 1 for the aldehyde includes a pentavalent carbon in the carbonyl unit—a highly unlikely species.) We conclude that the data do not provide evidence for the homogeneous distribution of CO in the 'copolymer' or for ketone structures. It is more likely that the structures of equation (1) or (2) represent the dominant form of the material.

The existence of chain distortions (to form solitons, polarons and bipolarons) associated with the charge storage states in conductive polymers has been proven unambiguously through the observation of doping-induced (or photo-induced) infrared modes. Although the soliton concept was initially developed for application to polymers with a degenerate ground state (such as *trans*-(CH)_x), these ideas have been successfully generalized (polarons and bipolarons) and applied to examples of the large class of conjugated polymers with non-degenerate ground states (for a review, see ref. 6 and references therein). The discovery that bond-length distortions always accompany charge-transfer redox reactions in such polymers in no way constrains the search for new conductive polymers. On the contrary, such insights have already begun to help in the development of a systematic

understanding of this growing class of materials.

F. WUDL
A. J. HEEGER

Department of Physics,
University of California,
Santa Barbara,
California 93106, USA

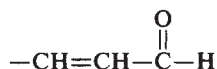
J. E. FROMMER

Allied Corporation,
Morristown,
New Jersey 07960, USA

- Chien, J. C. W., Babu, G. N. & Hirsch, J. A. *Nature* **314**, 723–724 (1985).
- Chien, J. C. W. & Babu, G. N. *J. chem. Phys.* **82**, 441–456 (1985).
- Chien, J. C. W. & Babu, G. N. *Macromolecules* **18**, 622–627 (1985).
- Schaefer, J. & Stejskal, E. *Topics in ^{13}C NMR Spectroscopy* Vol. 3 (ed. Levy, G. C.) (Wiley, New York, 1979) 297–315.
- Knothe, L. & Prinzbach, H. *Justus Liebigs Annl. Chem.* **687** (1977).
- Hieeger, A. J. *Polymer J.* **17**, 201 (1985).

CHIEN REPLIES—Wudl *et al.* comment on the nature of the copolymers described in our paper¹ and on the need to invoke the soliton hypothesis. They point out that the unpaired spin concentrations given in Table 2 of our paper are in disagreement with published results; this is caused by a typographical error as well as to the usage of repeat units for copolymers, which is unfamiliar to physicists. The heading for column 4 of Table 2 should read: "Spin per repeat unit ($\times 10^7$)" instead of ($+10^5$). The correct values are given in the paper by Chien and Babu². Whereas the spin concentration in $[\text{CH}]_x$ is customarily given in units of 'per CH', this is not possible for the copolymer: for example, the average repeat unit for copolymer C is $[(\text{C}_2\text{H}_2)_1(\text{CO})_{0.15}]_x$, as given elsewhere². For consistency, the spin concentration in homo-polyacetylene is given in terms of per C_2H_2 as the repeat unit. Thus the range of spin concentrations quoted for *trans*- $[\text{CH}]_x$ is 6.7×10^{-4} to 2×10^{-3} spin per repeat unit, or 1,500–500 repeat units per spin, or the more often cited 1,000–3,000 CH per spin. Therefore, the copolymers do indeed have very much lower spin concentrations than $[\text{CH}]_x$, as stated^{1,2}.

Another error in our paper is in the structure of unsaturated aldehyde, which should read



The doubts expressed by Wudl *et al.* concerning the homogeneous distribution of CO throughout the polyacetylene are unfounded. The structures of the materials have been characterized by most of the available and applicable techniques. Detailed kinetic studies^{2,3} exclude the action of CO as the chain-transfer or chain-termination agent and rule out the formation of block copolymers, alternat-

ing copolymers or a mixture of acetylene-CO copolymer and homo-polyacetylene. All the characterization results^{1,4} are in agreement with the copolymer having a random statistical distribution of CO. We did not—nor would any polymer chemist—assume a regular and discrete length for the polyene segments. For example, $[(\text{C}_2\text{H}_2)_1(\text{CO})_{0.15}]_x$ would have a number average acetylene sequence length of ~ 7 . The magic-angle spinning results^{1,4} have been confirmed using our new IBM NR-200AF spectrometer equipped with a sapphire specimen rotor: the infrared bands are not weak but broad, as expected for the presence of chemically distinct carbonyl groups.

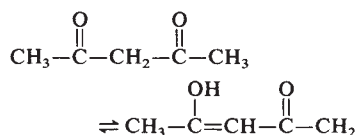
It is not our *basic premise* (attributed erroneously to us by Wudl *et al.*) that the $-\text{C}(=\text{O})$ groups prevent the unpaired spins (solitons) from moving. Quoting from our paper¹: "The EPR of *trans*-AC-copolymers shows about the same motional narrowing as in *trans*- $[\text{CH}]_x$. This is either because local processes cause line narrowing or because the $-\text{C}(=\text{O})$ — unit does not present a significant barrier to moving domain". We specifically did not postulate that the $-\text{C}(=\text{O})$ — groups prevent the unpaired spins from moving. Motional narrowing need not involve a long segment of conjugated backbone; all it requires is that the frequency of the motion is much greater than the hyperfine coupling constant. For example, the trimer of *p*-phenylene vinylene has a lightly doped ESR linewidth of 0.5 G (ref. 5). The same was observed for its low-relative molecular mass (M_r) analogues⁶. Also, *trans*- $[\text{CH}]_x$ with a M_r of 400 has an ESR linewidth comparable to that of the polymer prepared by the Shirakawa method ($M_r = 11,000$) and to that of *trans*- $[\text{CH}]_x$ having a M_r of 870,000 (ref. 7).

Wudl *et al.* concluded incorrectly that as the spin concentration is of the same magnitude for both the copolymers and *trans*- $[\text{CH}]_x$ and because they have similar thermopower, our experiments do not provide any true test of the soliton model. The fact that the undoped *trans*-AC-copolymers have spin concentrations $\sim 10^{-4}$ of that in *trans*- $[\text{CH}]_x$ but have thermopower comparable to that of *trans*- $[\text{CH}]_x$ showed that our experiments do provide a true test of the soliton model.

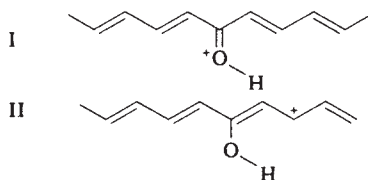
The properties of acetylene-CO copolymers are consonant with those of partially deuterated $[-(\text{CH}=\text{CH})_n\text{CHD}]_x$ materials⁸. There is also a fairly widespread view, based on anomalous Raman scattering results, that the conjugation length in *trans*- $[\text{CH}]_x$ ranges from 5 to 40. Wudl *et al.* infer from the high conductivity of the doped copolymer that the inclusion of $-\text{C}(=\text{O})$ — does not inhibit inter-chain transport, which is in fact the mechanism favoured by us².

Finally, Wudl *et al.* propose the existence of keto-enol tautomerization

[their reaction (1)] for the copolymer. But in fact, tautomerization is an internal rearrangement as exemplified by



Wudl *et al.* propose the formation of the conjugate acid of the copolymer, which we have shown would occur on treatment with acidic methanol. The thermodynamically stable form of the conjugate acid is I below:



Resonance forms such as II or others with the charge further removed, are less stable. Neither I nor II is an enol. Therefore, the argument of Wudl *et al.* that insertion of a CO-derived moiety need not interrupt conjugation does not merit consideration.

JAMES C. W. CHIEN

Department of Polymer Science
and Engineering,
Department of Chemistry,
University of Massachusetts,
Amherst, Massachusetts 01003, USA

- Chien, J. C. W., Babu, G. N. & Hirsch, J. A. *Nature* **314**, 723-724 (1985).
- Chien, J. C. W. & Babu, G. N. *J. chem. Phys.* **82**, 441-456 (1985).
- Schen, M. A., Karasz, F. E. & Chien, J. C. W. *J. Polym. Sci. Polym. Ed.* **21**, 2787-2812 (1983).
- Chien, J. C. W. & Babu, G. N. *Macromolecules* **18**, 622-627 (1985).
- Wnek, G. E., Chien, J. C. W., Karasz, F. E. & Lillya, C. P. *Polymer* **20**, 1441-1443 (1979).
- Gourley, K. D., Lillya, C. P., Reynolds, J. R. & Chien, J. C. W. *Macromolecules* **17**, 1025-1033 (1984).
- Chien, J. C. W. & Schen, M. A. *Macromolecules* (submitted).
- Yaniger, S. I., Kletter, M. J. & MacDiarmid, H. G. *188th Am. chem. Soc. Abstr. Polym.*, 154 (Publisher? Philadelphia, 1984).
- Wnek, G. E. *et al.* in *Conductive Polymers* (ed. Seymour, R. B.) 183-208 (Plenum, New York, 1981).

A quantum potential approach to the Wheeler delayed-choice experiment

ACCORDING to Bohm *et al.*¹, when a neutron is incident along SA of a neutron interferometer (Fig. 1), the wave-packet associated with the neutron splits into two wave-packets, one taking the path AEF and the other ABF, the neutron being in one of them. Although this approach satisfies the need that when a part of a wave passes along AEF, another part of the wave should pass along ABF to lead to interference at F, from the results already obtained from neutron interference experiments, it can be shown that their approach is as inadequate as the conventional approach.

Consider the case of a magnetic field H placed over a part of the path EF. If the wave-packet containing the neutron takes the path AEF, the contribution to the phase difference at F due to the field H depends on the mass m , the magnetic moment μ and the spin s of the neutron in the wave-packet. On the other hand, when the empty wave-packet takes the path ABF, to explain consistently the relative count-rates in the counters C_1 and C_2 , we have to accept that the contribution of

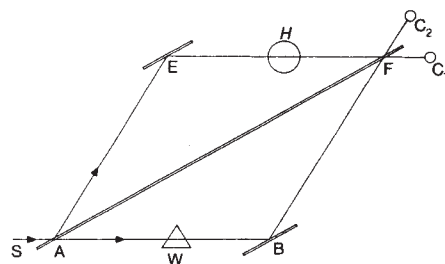


Fig. 1 Schematic arrangement of the neutron interferometer. AF, B and E are the reflecting planes, SA the direction of the incident neutrons, H a transverse magnetic field, W a material wedge and C_1 and C_2 are neutron counters.

the field H to the path difference at F is the same as before. As the neutron taking the path ABF cannot obviously be affected by the field H in the path AEF, this means that the empty wave-packet interacts with H in such a way as to lead ultimately to the same phase difference. In other words, the empty wave-packet is, in effect, not empty but has mass m , magnetic moment μ and spin s (as if it is the neutron).

Next, let us consider the case of H placed in the path EF and a material wedge W placed in the path AB. The phase shift produced by the wedge depends on the mass m and the short-range nuclear potential V of the neutron; we may regard this potential field V as moving with the neutron (just as its m , μ and s). It is well established that the experimentally observed relative count-rates of C_1 and C_2 can be attributed to the phase shifts due to H and W . There is no dispute about the way these phase shifts are estimated and related to the observed count-rates.

What has not been duly appreciated, either in the Copenhagen approach or in the quantum potential approach, is the implication of these estimates. In estimating the phase shift due to H , we accept that the mass m , the magnetic moment μ and the spin s of the incident neutron passes through H , and in estimating the phase shift caused by W we accept that the mass m and the potential field V of the same neutron passes through W . Evidently, some intrinsic property of the neutron (such as μ) cannot take one path (AEF), some other (such as V) the other path (ABF), and yet another (such as m) both paths. This is the basic enigma (or, should we say, the conceptual incon-

sistency) of quantum mechanics, yet to be resolved. In quantum potential approach, it is implied that in the neutron interferometer the empty wave-packet interacts with H and W as if it possesses all the properties of the neutron, without being a neutron. The enigma has assumed a new form.

N. UMAKANTHA

Department of Physics,
Karnatak University,
Dharwad 580 003, India

- Bohm, D. J., Dewdney, C. & Hiley, B. H. *Nature* **315**, 294-297 (1985).

BOHM ET AL. REPLY—We feel that Umakantha has not understood our proposal. He argues that because the empty wave-packet interacts with the magnetic field it must have all the properties of the particle, that is, mass m , magnetic moment μ and spin s . Our suggestion is that the neutron is a particle that passes along one path or the other. After it has entered one of the paths, it may still be affected by the wave-packet coming from the other. Even though the wave-packet that is 'empty' is determined by Schrödinger's equation, which contains the parameters m , μ and s , the presence of these parameters in no way implies that the corresponding physical properties have passed through along with the empty packet.

To make this point still more clear, we may use the notion that the wavefunction contains something analogous to information, along the lines that we have suggested elsewhere¹. Although the particle follows one path, its behaviour is still affected by information coming from the other path. The parameters m , μ and s appearing in Schrödinger's equation merely affect the propagation of the wavefunction and clearly do not imply that anything other than information has passed along the path not containing the particle.

An analogous case would be to have a set of computers programmed with these parameters which would correspondingly affect the behaviour of some external system. One would not want to say that, for example, mass, magnetic moment and spin were contained in the computers.

A similar discussion can be carried through for the second example of a wedge W placed in one of the paths.

In conclusion, it is in fact implied in the quantum potential approach that the 'empty wave-packet' is affected by the magnetic field, but it is clear that it does not therefore possess 'all the properties of the neutron'.

D. J. BOHM
C. DEWDNEY
B. H. HILEY

Department of Physics,
Birkbeck College,
Malet Street, London WC1E 7HX, UK

- Bohm, D. & Hiley B. J. *Fdn Phys.* **14**, 255-274 (1984).

Physics

An Axiomatic Basis for Quantum Mechanics. Vol. 1. Derivation of Hilbert Space Structure. By GÜNTHER LUDWIG. Springer-Verlag: 1985. Pp.243. ISBN 3-540-13773-4/0-387-13773-4. DM 118.

Elements of Quantum Mechanics of Infinite Systems. By F. STROCCHI. World Scientific: 1985. Pp.179. Pbk ISBN 9971-978-92-X. Pbk £16.95.

High Resolution in Solar Physics. R. MULLER (ed.). Springer-Verlag: 1985. Pp.320. ISBN 3-540-15678-X. Pbk DM 45.

Initiation and Growth of Explosion in Liquids and Solids. By F.P. BOWDEN and Y.D. YOFFE. Cambridge University Press: 1985. Pp.104. Pbk ISBN 0-521-31233-7. £7.95, \$14.95.

The Mechanical Universe: Introduction to Mechanics and Heat. By RICHARD P. OLENICK *et al.* Cambridge University Press: 1985. Pp.584. ISBN 0-521-30429-6. £17.50, \$24.95.

Practical Physics (3rd Edition). By G.L. SQUIRES. Cambridge University Press: 1985. Pp.215. Hbk ISBN 0-521-24952-X; pbk ISBN 0-521-27095-2. Hbk £22.50, \$44.50; pbk £7.95, \$17.95.

Shelter Island II. Proceedings of the 1983 Shelter Island Conference on Quantum Field Theory and the Fundamental Problems of Physics. ROMAN JACKIW *et al.* (eds). MIT Press: 1985. Pp.369. ISBN 0-262-10031-2. \$37.50.

Spin Fluctuations in Itinerant Electron Magnetism. By TORU MORIYA. Springer-Verlag: 1985. Pp.239. Hbk ISBN 3-540-15422-1. DM 110.

Treatise on Heavy-Ion Science. Vol.5: High Energy Atomic Physics. D. ALLAN BROMLEY (ed.). Plenum: 1985. Pp.498. Hbk ISBN 0-306-41575-5. Np.

Earth Sciences

Fluid Inclusion Studies. T. SHEPHERD, A.H. RANKIN and D.H.M. ALDERTON. Blackie: 1985. Pp.239. ISBN 0-412-00601-4. £26.

Models of the Atmosphere and Ionosphere. K. RAWER, C.M. MINNIS, K.S.W. CHAMPION and M. ROEMER (eds). Pergamon Press: 1985. Pp.234. Pbk ISBN 0-08-033196-3. £31.00, \$49.50.

Principles of Igneous Petrology. By SVEN MAALØE. Springer Verlag: 1985. Pp.374. Hbk ISBN 3-540-13520-0. DM 138.

Space Observations for Climate Studies. G. OHRING and H.J. BOLLE (eds). Pergamon Press: 1985. Pp.396. Pbk ISBN 0-08-033195-5. £31.00, \$49.50.

Biological Sciences

The Acute Stroke. By VLADIMIR HACHINSKI and JOHN W. NORRIS. F.A. Davis Company, Philadelphia, USA: 1985. Pp.286. Hbk ISBN 0-8036-4502-3. \$45.00.

Atlas of Prostatic Cytology. By W. LEISTEN-SCHNEIDER and R. NAGEL. Springer-Verlag: 1985. Pp.227. ISBN 3-540-13954-0. DM 220.

The Audubon Society Guide to Attracting Birds. By STEPHEN W. KRESS. Charles Scribner's Sons, USA: 1985. Pp. 377. Hbk ISBN 0-684-18362-5. \$24.95.

Biology of Deer Production. P.F. FENNESSY and K.R. DREW (eds). Royal Society of New Zealand, Bulletin 22: 1985. Pp.482. Pbk ISBN 0-908654-04-9. Np.

The Biophysical Basis of Excitability. By HUGO GIL FERREIRA and MICHAEL W. MARSHALL. Cambridge University Press: 1985. Pp.484. Hbk ISBN 0-521-30151-3. £50.00, \$80.00.

Branching Processes Applied to Cell Surface Aggregation Phenomena. By CATHERINE A. MACKEN and ALAN S. PERELSON. Springer-Verlag: 1985. Pp.122. Pbk ISBN 3-540-15656-9. DM 25.

Case Studies in Population Biology. L.M. COOK (ed.). Manchester University Press: 1985. Pp.218. ISBN 0-7190-1740-8. £22.50.

Comparative Genetic Toxicology. J.M. PARRY and C.F. ARLETT (eds). Macmillan Press: 1985. Pp.640. Hbk ISBN 0-333-38897-6. £60.00.

Complex Arrhythmias: Self Assessment. By EDWARD K. CHUNG. Karger: 1985. Pp.310. Hbk ISBN 3-8055-3639-9. SwFr 82, DM 98, £25.70, \$35.00.

Craniofacial Dysmorphism: Studies in Honour of Samuel Pruzansky. M. MICHAEL COHEN Jr. and BEVERLY R. ROLLNICK (eds). Alan R. Liss: 1985. Pp.364. Hbk ISBN 0-8451-0247-8. £58.00.

Current Topics in Bioenergetics. C.P. LEE (ed.). Academic: 1985. Pp.192. ISBN 0-12-152514-7. \$39.50, £35.

Dolphin Dreamtime: Talking to the Animals. By JIM NOLLMAN. Muller, Blond and White: 1985. Pp.167. ISBN 0-85634-199-1. £10.95.

An Ecological and Behavioural Study of the Pig-Tailed Macaque. By J.L. CALDECOTT. Karger: 1985. Pp.262. Hbk ISBN 3-8055-4212-7. SwFr 116, DM 139, £36.30, \$49.50.

Endothelial Cell Vesicles. F. HAMMERSEN and D.H. LEWIS (eds). Karger: 1985. Pp.173. Pbk ISBN 3-8055-4070-1. DM 95, \$33.75, £24.70.

Evolution of Prokaryotes. KARL H. SCHLEIFER and ERKO STACKEBRANDT (eds). Academic: 1985. Pp.367. ISBN 0-12-626130-4. \$30, £27.

Fibrosis. Ciba Foundation Symposium 114. Pitman: 1985. Pp.256. Hbk ISBN 0-272-79814-2. £27.95.

A Field Manual of the Ferns and Fern-Allies of the United States and Canada. By DAVID B. LELLINGER. Smithsonian Institution Press: 1985. Pp.389. Hbk ISBN 0-87474-602-7. \$45.00.

Fish Endocrinology. By A.J. MATTY. Croom Helm: 1985. Pp.267. ISBN 0-88192-024-X. £27.95.

Immunology and Immunopathology of Reproduction. V. TODER and A.E. BEER (eds). Karger: 1985. Pp.212. Hbk ISBN 3-8055-4059-0. SwFr 156, DM 187, £48.80, \$66.50.

Information Processing in Biological Systems. BEHRAM KURSUNOGLU (chairman), STEPHEN L. MINTZ and ARNOLD PERLMUTTER (eds). Plenum: 1985. Hbk ISBN 0-306-42075-9. Np.

An Introduction to Coastal Ecology. By P.J.S. BOADEN and R. SEED. Blackie/Chapman & Hall: 1985. Pp.218. ISBN 0-412-01021-6. Hbk £19.95; pbk £9.95.

The Leopard's Tale. By JONATHAN SCOTT. Elm Tree Books, Garden House, 57-59 Long Acre, London WC2 9JZ, England: 1985. Pp.192. ISBN 0-241-11444-6. £15.

Molecular Architecture of Proteins and Enzymes. RALPH A. BRADSHAW and JORDAN TANG (eds). Academic: 1985. Pp.323. Hbk ISBN 0-12-124570-5. £43.50, \$50.00.

Molecular Biology of Aging. AVRIL D. WOODHEAD, ANTHONY D. BLACKETT and ALEXANDER HOLLAENDER (eds). Plenum: 1985. Pp.482. Hbk ISBN 0-306-42084-8. Np.

Minerals in Food and Nutritional Topics. G.H. BOURNE (ed.). Karger: 1985. Pp.258. ISBN 3-8055-4058-2. SwFr 220, \$93.75.

Progress in Nickel Toxicology. STANLEY S. BROWN and F. WILLIAM SUNDERMAN Jr. (eds). Blackwell Scientific: 1985. Pp.250. Hbk ISBN 0-632-01355-9. £27.50.

Protein Transport and Secretion. Current Communications in Molecular Biology. MARY-JANE GETHING (ed.). Cold Spring Harbor Laboratory, New York, USA: 1985. Pp.215. Pbk ISBN 0-87969-183-2. \$30.00.

Pulmonary Development: Transition from Intra-uterine to Extrauterine Life. G.H. NELSON (ed.). Dekker: 1985. Pp. 526. ISBN 0-8247-7316-0. £75.

Review of Biological Research in Aging. Vol 2. MORTON ROTHSTEIN (ed.). Alan R. Liss: 1985. Pp.561. Hbk ISBN 0-8451-3501-5. £73.00.

Substance P: Metabolism and Biological Actions. C.C. JORDAN and P. OEHME (eds). Taylor & Francis: 1985. Pp.260. ISBN 0-85066-324-5. £30.

Systematics and Biogeography of the Longicorn Beetle Tribe Tmesisternini. By J. LINSLEY GRESITT. Pacific Insects Monograph, Department of Entomology, BP Bishop Museum, Honolulu, Hawaii: 1984. Pp.263. Pbk \$23.50.

The World of Biology. By THOMAS G. OVERMIRE. Wiley: 1985. Pp.579. Hbk ISBN 0-471-88000-0. £25.50.

World's Debt to Pasteur. HILARY KOPROWSKI and STANLEY A. PLOTKIN (eds). Alan R. Liss: 1985. Pp.342. ISBN 0-8451-2002-6. £30.

General

Advances in Information Systems Science, Vol 9. JULIUS T. TOU (ed.). Plenum: 1985. Pp.340. Hbk ISBN 0-306-41644-1. Np.

Alias Papa: A Life of Fritz Schumacher. By BARBARA WOOD. Oxford University Press: 1985. Pp.394. ISBN 0-19-281881-3. £4.95.

Architecture of Systems Problem Solving. By GEORGE J. KLIR. Plenum: 1985. Pp.539. ISBN 0-306-41867-3. Np.

Arctic Underwater Operations. LOUIS REY (eds). Graham & Trotman: 1985. Pp.364. ISBN 0-860-10-6314. \$69, £43.

Autolization: Recent Developments and Applications. AARON TEMKIN (ed.). Plenum: 1985. Pp.261. Hbk ISBN 0-306-41854-1. Np.

Basic Biogeography, 2nd Edn. By NIGEL PEARS. Longman: 1985. Pp.358. Pbk ISBN 0-582-30120-3. Pbk £8.95.

Bioactive Polymeric Systems: An Overview. CHARLES G. GEBELEIN and CHARLES E. CARRAHER Jr. (eds). Plenum: 1985. Pp.689. Hbk ISBN 0-306-41855-X. Np.

Capitalism and Leisure Theory. By CHRIS ROJEK. Tavistock Publications: 1985. Pp.208. Hbk ISBN 0-422-79060-5. £14.00.

The Changing Year: Proteus of The Field. By WILSON STEPHENS. Unwin: 1985. Pp.182. ISBN 0-04-910084-X. £2.95.

Earth and Life Through Time. By STEVEN M. STANLEY. W.H. Freeman: 1985. Pp.671. ISBN 0-7167-1677-1. \$33.95.

Exploring Data Tables, Trends, and Shapes. DAVID C. HOAGLIN, FREDERICK MOSTELLER and JOHN W. TUKEY (eds). Wiley: 1985. Pp.527. Hbk ISBN 0-471-09776-4. £40.85.

Expository Science: Forms and Functions of Popularisation. TERRY SHINN and RICHARD WHITLEY (eds). Reidel: 1985. Pp.292. Hbk ISBN 90-277-1831-8; pbk ISBN 90-277-1832-6. Dfl. 68, \$24, £18.95.

Extraterrestrials: Science and Alien Intelligence. EDWARD REGIS Jr. (ed.). Cambridge University Press: 1985. Pp.278. ISBN 0-521-26227-5. £25, \$39.50.

Free Rings and their Relations (second edition). By P.M. COHN. Academic: 1985. Pp.588. Hbk ISBN 0-12-179152-1. \$80.00, £66.50.

Global Change. T.F. MALONE and J.G. ROEDERER (eds). Cambridge University Press: 1985. Pp.512. ISBN 0-521-30045-2. \$19.95, £12.95.

Handbook of Laboratory Waste Disposal. By MARTIN J. PITT and EVA PITT. Wiley: 1985. Pp.360. ISBN 0-85312-634-8. £39.95.

Kuwait City Parks: A Critical Review of Their Design Facilities Programs and Management. By SUBHI ABDULLAH AL-MUTAWA. KPI, 14 Leicester Square, London WC2H 7PH, England: 1985. Pp.125. ISBN 0-7103-0068-9. £20.

Land into Landscape. By JOHN MICHAEL HUNTER. Longman/George Godwin: 1985. Pp.182. ISBN 0-7114-5625-9. £25.

Living with the California Coast. GARY GRIGGS and LAURET SAVOY (eds). Duke University Press: 1985. Pp.393. Hbk ISBN 0-8223-0632-8; pbk ISBN 0-8223-0633-6. Hbk \$27.95; pbk \$14.95.

The Living State — II. R.K. MISHRA (ed.). Wiley: 1985. Pp. 699. ISBN 9971-978-26-1. £68.75.

Love and Tradition: Marriage Between Jews and Christians. By EGON MAYER. Plenum: 1985. Pp.311. Hbk ISBN 0-306-42043-0. Hbk \$17.95.

Modern Science and Human Values. By W.W. LOWRANCE. Oxford University Press: 1985. Pp.250. Hbk ISBN 0-19-503605-0. £25.00.

The Natural Alien: Humankind and Environment. By NEIL EVERNDEN. University of Toronto Press: 1985. Pp.160. ISBN 0-8020-2552-8. \$19.95.

The New Story of Science. By ROBERT M. AUGROS and GEORGE N. STANCIU. Gateway Edition, 940 North Shore Drive, Lake Bluff, IL 60044: 1985. Pp.234. Pbk ISBN 0-89526-833-7. Pbk \$6.95.

Policies for Common Security. STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE. Taylor & Francis: 1985. Pp.250. ISBN 0-85066-301-6. £16.

Realism and the Aim of Science. W.W. BARTLEY III (ed.). Hutchinson: 1985. Pp.240. Pbk ISBN 0-09-151451-7. Pbk £11.95.

Pulsed-field gel electrophoresis of large DNA molecules

from Cassandra L. Smith and Charles R. Cantor

A new development in gel electrophoresis allows investigators to fractionate and analyse DNA species from 10,000 to more than a million base pairs.

In conventional electrophoresis, using a constant electrical field, DNA molecules behave as free draining species. In free solution all linear DNA molecules have the same mobility, independent of size. In a gel, high resolution size fractionation can be achieved because of the sieving properties of the gel matrix. Directed by the electrical field, small DNA molecules will move through the gel in a relatively straight path, while very large DNA must take a more indirect path to find gel pores large enough to accommodate it.

The sieving properties of a gel become irrelevant once a molecule is larger than the largest gel pores. For DNA in agarose, sieving properties are lost when the size exceeds about 20,000 base pairs. Under these circumstances, most molecules act like rigid objects and will not enter the gel at all. However, the stiff DNA coils become distorted under the influence of the electrical field, and pass through the network of gel pores by movement along their long axes. Once this process, called reptation, occurs, all molecules move with the same velocity because all travel through the same pore network.

Altered shapes

The technique of pulsed-field gel (PFG) electrophoresis was designed to circumvent this loss of resolution and extend the power of gel electrophoresis to handle larger DNA molecules¹. In PFG electrophoresis the separation of molecules is based on the rate at which these molecules

alter their shape inside the gel matrix². A number of variations of the technique have already been developed³⁻⁵. Here we describe the most common method.

Apparatus for PFG electrophoresis is shown schematically in Fig. 1. It consists of a horizontal submarine gel box containing two independent perpendicular arrays of electrodes, termed north (N), south (S), east (E) and west (W). The 1 to 1.5 per cent agarose gel, typically a 20 cm square, is placed at 45 degrees in the centre of the apparatus. Electrical power is switched alternately between the N-S and E-W electrode arrays. Depending on the size range of the DNA molecules, each electrical pulse continues for one second to more than five minutes. DNA molecules exposed to the alternating fields first migrate 'south', then 'east'. Each time the field direction changes, the molecules must spend part of the next pulse time relaxing (changing shape and orientation) before they attain configurations that allow efficient net translation through the gel.

The particular electrode geometries shown in Fig. 1 serve several important functions. Because the sum of the two alternate fields is symmetric about the diagonal of the apparatus, the net motion of the DNA molecules is a relatively straight path along the diagonal. The field shapes lead to angles between alternate fields that range from 100° to 150°. These large angles appear to enhance resolution. Because the net field along the diagonal of the gel decreases with distance from the starting point, each band of DNA molecules self-sharpens, further enhancing the resolution of the technique⁴.

Relaxation

While the mechanism of PFG electrophoresis remains unproven, the following working hypothesis is consistent with available results. DNA relaxation times are roughly proportional to DNA size. As relaxation times approach the pulse duration, the net mobility varies very sharply with molecular weight and very high electrophoretic resolution is achieved. Once relaxation times become longer than the pulse time, the DNA molecules stop responding to the alternating fields and their shape in the gel becomes time independent; they all move with the same velocity, and no size fractionation occurs. Thus the choice of a particular pulse time sets an

upper limit to the DNA size range amenable to analysis. A typical set of PFG electrophoresis results is shown in Fig. 2. The behaviour of λ vir DNA concatamers, which differ in size by a constant increment of 42,500 base pairs, illustrates the high resolving power of the method from the 42,500 base pair λ vir monomer up to the 1,190,000 base pair 28-mer.

The power of PFG electrophoresis can be realized only if large DNA molecules can be prepared in discrete sizes. One popular method involves suspending cells from virtually any source in liquid low-gelling agarose and allowing the mixture

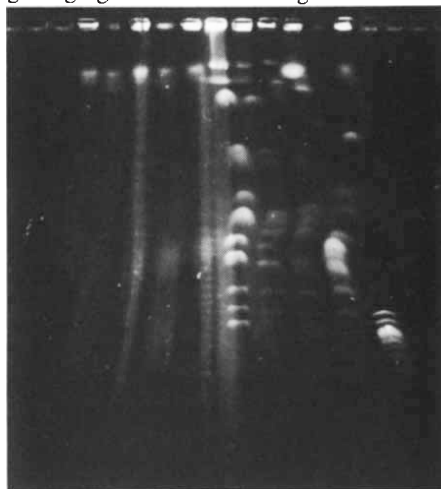


Fig. 2 Typical PFG data. The sample wells are shown at the top of an ethidium-stained gel. Lanes 1,3,5 and 11 are λ vir DNA concatamers used as length standards; lanes 6-8, and 9 are chromosomal DNA from three different yeast strains and one *Leishmania* species, respectively; lanes 2 and 4 are a digest of mouse DNA with the enzyme *NotI*, while lane 10 is a similar digest of *E. coli* DNA.

to solidify in a mould¹. The resulting agarose block, typically 2 mm by 5 mm by 10 mm, is then treated with a variety of enzymes and detergents to lyse the cells within, removing all protein and RNA. Since the procedure is essentially quantitative, the appropriate cell number can be calculated from the genome size, to arrive at final DNA contents of 0.1 to 5 μ g. These are effective DNA concentrations for high resolution PFG electrophoretic separations¹.

Cleave and run

Organisms like yeasts^{2,3,5}, and various unicellular parasites⁶⁻⁸, have intact chromosomal DNA that ranges in size from

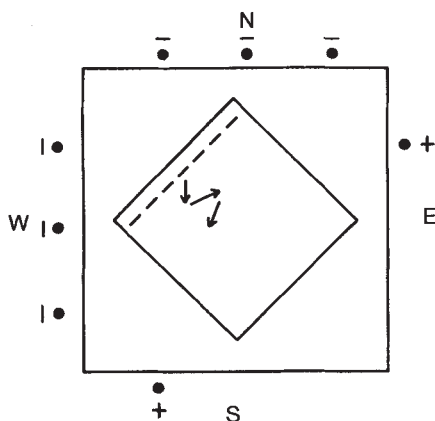


Fig. 1 Schematic illustration of typical PFG electrophoresis apparatus. The horizontal and vertical electrodes are shown as sets of points. The gel is placed at 45° and arrows indicate the direction of motion of DNA molecules starting from the central sample well.

100,000 base pairs to a few million base pairs. DNA from these organisms can be analysed directly by simply placing the agarose sample block into a well cut in the running gel and applying the appropriate pulsed fields. The results are separated chromosomes as shown in Fig. 2. With these results, a variety of cytogenetic procedures can be accomplished directly by electrophoresis. For example, cloned probes can be assigned to chromosomes by ordinary Southern blot analysis; translocations, deletions and insertions can be visualized by direct changes in DNA size².

Bacteria usually contain single circular DNA molecules with sizes ranging from less than one million base pairs to more than ten million. To analyse such genomes, DNA prepared in agarose blocks is cleaved *in situ* by treating the blocks with restriction endonucleases⁴. Particularly convenient for many organisms are the enzymes *NorI* and *SfiI*, which have 8 base recognition sequences and make large DNA fragments. For *Escherichia coli* such fragments average a few hundred thousand base pairs, so all fall into the separation range of PFG electrophoresis. Gene mapping and the analysis of genomic rearrangements can be carried out as described earlier for yeast.

A typical mammalian chromosome contains a single DNA molecule of about 150 million base pairs. The appropriate choice of restriction nucleases with sites rare in particular genomes generates a variety of discrete, easily separated DNA pieces. With the appropriate digestion, PFG electrophoresis will analyse mammalian chromosomes as easily as bacterial DNA⁹.

Many procedures for the analysis or manipulation of DNA become easier and more sensitive in direct proportion to DNA size. Thus PFG electrophoresis appears to have broad applicability to a wide range of problems in molecular and cell biology. The relative simplicity of PFG electrophoresis, and the imminent availability of commercial equipment for the technique (from LKB-Producter AB), should make this technology readily available to a large number of laboratories. □

Cassandra L. Smith and Charles R. Cantor are at the Department of Genetics and Development, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA.

1. Schwartz, D.C. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **47**, 189-195 (1983).
2. Schwartz, D.C. & Cantor, C.R. *Cell* **37**, 67-75 (1984).
3. Carle, G.F. & Olson, M.V. *Nucleic Acids Res.* **12**, 5647-5664 (1984).
4. Smith, C.L., Warburton, P.W., Gaal, A. & Cantor, C.R. in *Genetic Engineering 8* (eds Setlow, J.K. & Hollaender, A.) (Plenum, New York, in the press).
5. Carle, G.F. & Olson, M.V. *Proc. natn. Acad. Sci., U.S.A.* **82**, 3756-3760 (1985).
6. Van der Ploeg, L.H.T., Schwartz, D.C., Cantor, C.R. & Borst, P. *Cell* **37**, 77-84 (1984).
7. Spithill, T.W. & Samaras, N. *Nucleic Acid Res.* **13**, 4155-4169 (1985).
8. Kemp, D.J. *et al.* *Nature* **315**, 347-350 (1985).
9. Smith, C.L. *et al.* *Meth. Enzym.* (in the press).

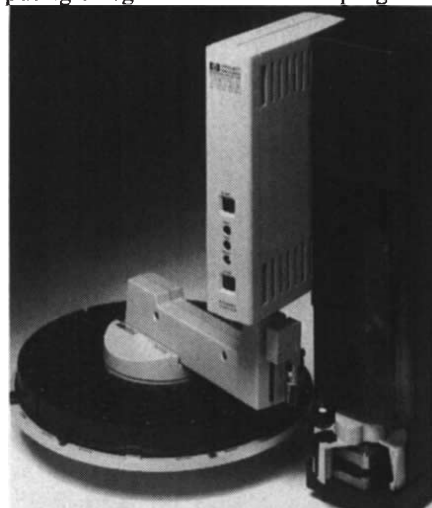
Currents in separation

Features on products, catalogues and courses bring a host of separation items together.

THIS spring gas chromatographers can get a special deal on Hewlett-Packard's latest **integrated chromatography system** (Reader Service No. 100). The HP 5890 gas chromatograph makes its debut on the UK market with a £1,500 discount on all orders placed before 30 April 1986. The complete system includes the HP 7673A automatic sampler for fast injection without heat transfer from the port to the syringe needle, and the HP 3393A computing integrator with BASIC program-

three times faster than packed column procedures. The columns, at \$245 to \$425 (US), fit most gas chromatographs.

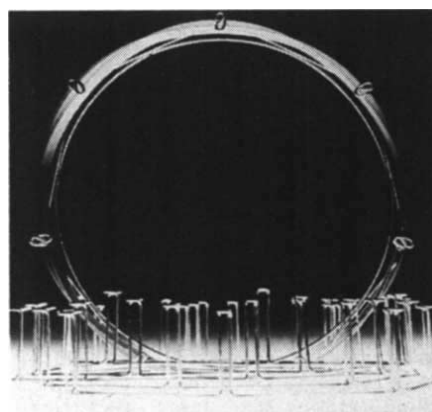
The model 93 **gas chromatograph** from Ai Scientific will handle routine analysis work at a price that's not so ordinary: less than £4,500 (UK) (Reader Service No. 102). Equipped with a large column oven and a programmable temperature control, the model 93 can accommodate packed, capillary or wide-bore capillary column.



The automatic sampler gives H-P's gas chromatograph a boost.

ming. In laboratories requiring high sample throughput and reproducibility, the automated GC package can hold its own.

J. & W. Scientific, Inc. has gas chromatography columns all tied up. Their Megabore **silica capillary columns** come in 15 m and 30 m lengths wound to manageable sizes (Reader Service No. 101). Thin, flexible Megabores separate without conventional packing, which tends to dampen

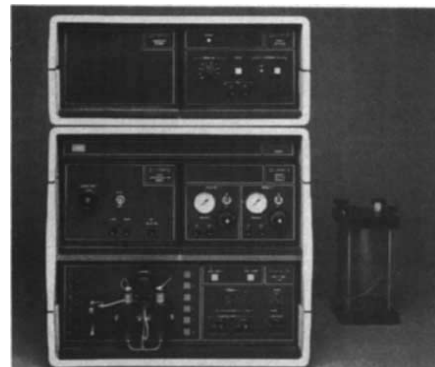


J & W Scientific has columns with a twist.

pressure, and eliminate surface pretreatment, silanization and rejuvenation. Conditioning takes 15 minutes; separations go

High performers

Two ion chromatograph accessories from Dionex allow both **monovalent and divalent cation analysis** to be performed in the same run (Reader Service No. 103). The HPIC-CS3 cation suppressor column is geared to the individual or simultaneous analysis of alkali metals and alkaline



Accessories for the 2000i are sold individually or as a package.

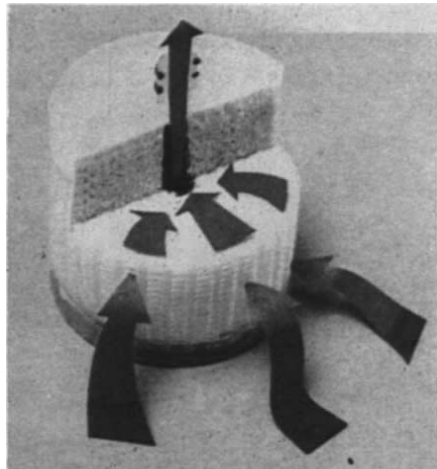
earth. The cation micromembrane suppressor (CMMS) uses ultra-thin membranes and ion exchange screens to complete the analysis. The component combination, priced at \$1,265 (US), is designed to accompany Dionex's Series 2000i ion chromatographs.

Greater chiral recognition reflects the new generation of stationary phase materials from Technicol Ltd. (Reader Service No. 104). As part of the Cyclobond series of HPLC columns, the material begins with cyclodextrin molecules bound to a 5 µm silica gel. But the cyclodextrin molecules in Technicol's columns are derivatized to enhance hydrogen bonding between the rim of cyclodextrin and the guest molecule. A substituted beta-cyclodextrin, for instance, can achieve baseline resolution of DL-norgestrel.

These notes are compiled in the Nature office by Karen Wright, based on manufacturers' information. For further details about products mentioned here use the reader service card bound inside the journal.

Compounds with phenolic functional groups and certain alcohols and amines will also go their separate ways on Technicol's columns, which retail for about £300 (UK).

For protein purification, pre-packed ZetaPrep disks and cartridges deliver high

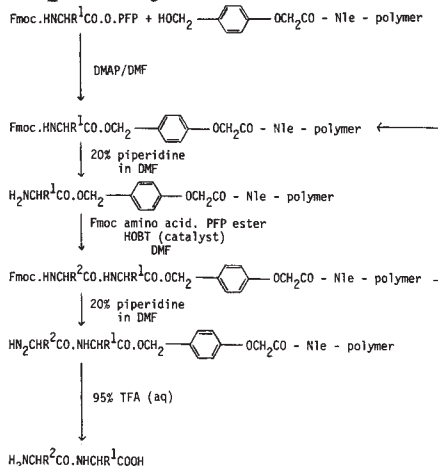


ZetaPrep disks can replace column and batch techniques.

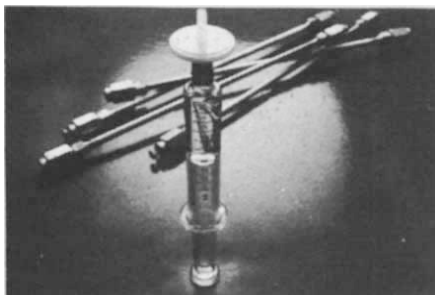
yields at flow rates approaching 25 litres per minute (Reader Service No. 105). Cuno Molecular Separations makes **ion exchangers** with DEAE, QAE and SP functionalities anchored to a solid cellulose-based matrix. This structure resists channelling and dimensional changes that can beset other ion exchange systems, at prices between \$17 per disk and \$150 per cartridge kit (US).

Millipore Corp. has expanded its Millex line of HPLC clarification filters to en-

Peptide synthesis: Erratum



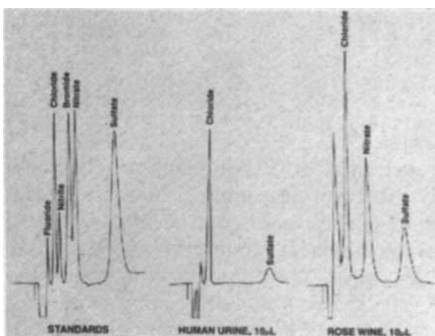
In R.P. Andrew's article on continuous flow peptide synthesis (*Nature* 319, 429; 1986), the protocol shown in Fig. 1 was the synthesis sequence for the Merrifield chemistry, not the Fmoc polyamide method. The correct polyamide synthesis, using Fmoc amino-acid PFP esters, is given above.



HPLC samples come clean with Millipore

compass 1 to 100 ml capacities (Reader Service No. 106). The Millipore design minimizes hold-up volume in units measuring 4 mm (less than 10 µl after air purge) and 25 mm (less than 100 µl hold-up volume). Prices vary according to filter features such as size, coarse or fine filtration and solvent and sample specifications.

'Indirect photometric detection' is a somewhat long-winded term for a technique that abbreviates **inorganic ion separation**. Kratos Analytical incorpo-



A new Kratos system turns ion chromatography inside out.

rates that technique in an HPLC-based system that removes the requirements for ion suppression and conductivity detection (Reader Service No. 107). Instead, Kratos's system uses an ultraviolet-absorbing compound in the mobile phase. Non-absorbing anions eluted from the column show up as negative peaks when an ultraviolet detector monitors the eluant. In the United States the system sells for \$8,995 to \$12,000.

With Fotodyne Inc's agarose gel system, what used to take a little walking now takes just two feet. The Foto/Phoresis I



Fotodyne's gel equipment is ideal for cramped quarters.

combines the apparatus for electrophoresis, visualization and photodocumentation in one DNA analysis package that occupies two square feet of bench space (Reader Service No. 108). The system includes a 250 V power supply, a hand-held instant format camera and an ultraviolet-blocking cover for less than \$1,600 (US). DNA banding reproduction is 1:1, providing an exact replica for future reference.

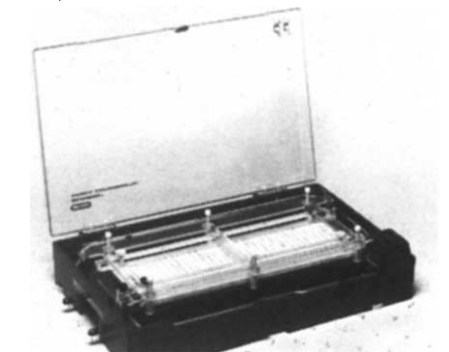
Going with the flow

The struggle to circumvent gravity's effects put electrophoresis on the space



Electrophoresis in low gravity from the Elphor VaP 21.

shuttle, but on Earth **free flow electrophoresis** offers an alternative solution. Protein Technologies, Inc. now markets the Elphor VaP 21, a device that performs field step and zone electrophoresis, isoelectric focusing and isotachopheresis using a continuous laminar flow of sample and carrier electrolyte (Reader Service No. 109). A liquid cooling system carries off heat, and fractions are separated and collected by a 90-tube array. Protein Technologies is introducing the VaP 21 in the United States at \$49,500, including installation, training and a one-year warranty.



The horizontal unit from Bio-Rad is an easy cell.

The **horizontal electrophoresis cell** from Bio-Rad may be ideal for analytical, ultra-thin, agarose or preparative electrofocusing, but deep down it's cold-hearted. At



Nelson's data system goes hand in hand with the Hewlett-Packard series 300 computer.

the centre of the Bio-Phoresis cell, a ceramic cooling platform draws heat from gels while a condensation coil keeps condensate under control (Reader Service No. 110). Several other features join a \$615 price-tag (US): adjustable platinum electrodes that exert uniform pressure; 375 ml buffer chambers for overnight runs; a permanent location grid on the cooling stage surface; and room enough to run 80 samples simultaneously.

Detector selections

In the United Kingdom, Applied Chromatography Systems Ltd. has launched an **electrochemical detector** to complement its pulse-free pumps. (Reader Service No. 111). The model 350/06 detector operates in oxidation or reduction modes, directed by an electronic control system. Its 0.5 μ l flow-cell incorporates a self-cleaning design that prolongs the life and upgrades the sensitivity of the 350/06.

A modular, **variable wavelength HPLC detector** from IBM screens out noise without touching the volume (Reader Service

No. 112). Its high-energy optics and electronics system spans ultraviolet and visible wavelengths from 190 to 700 nm, split into nine absorbance ranges. Starting at \$4,995



Flow assemblies switch in a jiffy with IBM's detector.

(US), the LC/9563 detector switches easily between assemblies for microbore, analytical, fast LC and semi-preparative applications. It can interface with the LC/9560 liquid chromatograph and join the LC/9505-SE automatic sample handler and the system 9000 chromatography software for full automation.

Diode array detection from Hewlett-Packard can generate up to 8 spectra per second, measuring ultraviolet and visible absorbance at wavelengths between 190 and 600 nm simultaneously. Combined with an HP 9000 series 300 computer workstation, the HP 1040M array detector can collect large amounts of LVC information for display or manipulation (Reader Service No. 113). A chromatogram can indicate the optimum for each eluting compound, and the touch of a sing-



Diode array detection brings H-P's workstation to its full potential.

le key will convert spectra into iso-absorbance data. The HP 1040 detector goes for £7,130 (UK); the HP 9000 series 300 costs about £12,000 with software.

In biomedical and pharmacological studies, electrochemical detectors like Gil-

son Medical Electronics Inc.'s model 141 can compile **HPLC profiles for electroactive samples** (Reader Service No. 114). The glassy carbon electrode monitors oxidation or reductive reactions occurring on its surface. Gilson claims sensitivity in the low femtomole range. Among the 141's more tangible advantages are user-selectable parameter presentation, an easy-access cell compartment and a built-



The electrochemical detector completes Gilson's HPLC monitoring line.

in test cell. The detector will be available early this spring with a price tag between \$3,500 and \$4,500 (US).

Data automata

Philips Analytical's gas chromatography systems are getting a new look with updated **graphics software** (Reader Service No. 115). A dual disc unit houses more than 120 methods to fulfil the analytical requirements of a multitude of samples;



Philips's software renders gas chromatography in graphic detail.

then, the graphics option will present up to three channels of data simultaneously. The program can alter parameters and assemble hard copies when necessary. The gas chromatography graphic option costs £1,600 (UK).

A **chromatography data system** from Nelson Analytical takes capillary, gas and liquid chromatography in its stride. The series 4430X system is equipped with a Motorola microprocessor, and the software can handle data from up to 12 chromatographs (Reader Service No. 116). Nelson's \$14,000 (US) data package includes routines for method and sequence generation, post-run reprocessing, reintegration, comparison, subtraction and ratios. Special applications software and full colour versions are also available for a personal touch.

ADVERTISEMENT

Xenopus Oocytes injection?

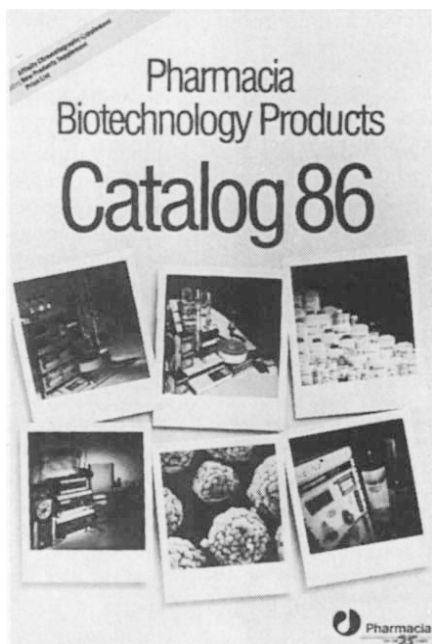


SINGER MKI MICROMANIPULATOR

We also offer a Microsyringe Kit.

Write to:

Singer Instrument Co. Ltd.
Treborough Lodge
Roadwater, Watchet
Somerset TA23 0QL
England. Tel. (0984) 40226
Telex: 337492 Comcab G



Pharmacia's best sellers.

Solvent delivery, injection and integration during HPLC analysis can suffer from human error, but Bio-Rad's **automation package** puts the process in safe, though inanimate, hands (Reader Service No. 117). The program also provides a results record and quantitative evaluation of peaks. The package sells for under \$14,000 (US), with a \$1,570 raw data option that allows reintegration and peak labelling.

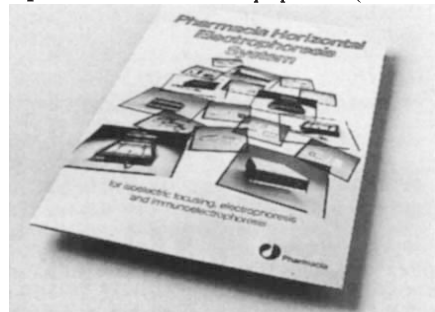
Personal computers have found a home in the chromatography laboratory, and Spectra-Physics offers the hardware along with the software. The **LABNET-XT/AT data system** does everything a personal computer can do, and then some: it stores raw chromatographic data and displays and reprocesses chromatograms through connections to Spectra-Physics computing integrators (Reader Service No. 118). The multi-tasking software can juggle remote data acquisition and reprocessing at the same time. A 10-megabyte system, which includes an IBM PC XT, is priced at \$10,000 (US), and the 30-megabyte IBM PC AT system costs \$12,000 (US).



LABNET captures the essentials in the chromatography laboratory.

Brochures in brief

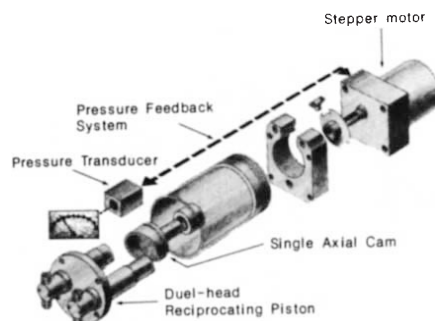
Pharmacia presents their **FBE-3000 horizontal electrophoresis system** with full colour and vivid detail in a new pamphlet (Reader Service No. 119). Instruments, chemicals and accessories to run major forms of IEF can be previewed in Pharmacia's publications. The booklet also outlines seven application kits which expand the system's capabilities for polyacrylamide IEF, immunoelectrophoresis and titration curves, among others. Pharmacia has also released their 1986 **catalogue of separation media** and equipment (Reader



Electrophoresis from beginning to end.

Service No. 120). Their chemicals, instruments and fittings line-up ranges from liquid chromatography and electrophoresis to cell separation.

Liquid chromatography components, enlivened with photographs, diagrams and chromatograms, are the main charac-



Varian illustrates its liquid chromatography equipment.

ters in Varian Associates, Inc.'s Slim-Line series 2000 brochure (Reader Service No.

ADVERTISEMENT

CompuGene makes DNA Sequencing Rapid & Accurate



Gather and analyze nucleotide data on your personal computer or VAX.

GelReader System

Acquisition of DNA sequence data is rapid enough that multiple readings may be taken and compared with a minimum of the scientist's energy.

Verbal Feedback

All GelReader program prompts and DNA input are spoken to the user and shown on the screen.

The Most Compatible

All CompuGene software runs on the IBM PC and AT compatibles, (including the Olivetti M24) and on the Dec VAX and MicroVAX.

DNALYSIS Software

Restriction analysis with scale map, ORF's, and site labels; genetic code translation and reverse translation; complete restriction search of the reverse translation; more.

Chromosome Gel Control

CompuGel program controls precise, complicated timing cycles for whole chromosome gels or other 110v equipment through the IBM PC type parallel port.

Contact

CompuGene

15 Carter Court

St. Louis, MO 63132

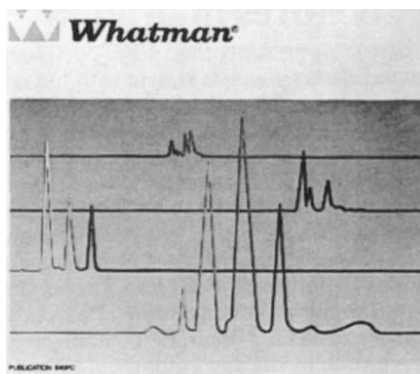
(314) 994-1587 or 531-6061

(800) 231-4545 (tone) G-E-N-E

121). The brochure describes performance features and specification for pumps, gradient programmers, detectors and peripheral equipment.

For the latest product information on **thin layer chromatography**, Camag compiled a TLC-86 catalogue replete with instrumentation and instructions for selection (Reader Service No. 122).

For tips on **preparative and process scale chromatography**, Whatman's 20-page guide to media and columns may come in handy (Reader Service No. 123). Unique specialized materials, silica-based HPLC media and anion and cation ex-



Materials and methods reviewed in Whatman's chromatography guide.

change celluloses get equal time in Whatman's book.

A 32-page guide from Manville Filtration and Minerals fills out a discussion of Chromosorb **diatomite supports** with information on basic chromatography theory and a glossary of terms (Reader Service No. 124). The literature gives physical, chemical and chromatographic properties for the five basic Chromosorb grades, as well as performance characteristics and surface treatments.

A lot to learn

Training programmes for fledgling separation specialists are available in slide/audio tape format and videotapes from Autoscribe Ltd. (Reader Service No. 124). Presentations cover gas, liquid and thin layer chromatography as well as electrophoresis, and costs vary from £160 to £540 (UK). HPLC principles and column selection, qualitative and quantitative analyses and troubleshooting the liquid chromatograph are a few of the areas dealt with in depth.

Chrompack Ltd. has organized 3-day **HPLC courses** for new and inveterate chromatographers alike (Reader Service No. 125). The basic course utilizes lectures and practical sessions to give beginners a foundation of theory, practice and manual skills. The advanced course requires previous experience and discusses the latest developments in HPLC. Both courses are scheduled for April in the UK; the basic session costs £210 and the advanced £250, plus VAT. □

Coming soon to Product Review

- **Pittsburgh Conference.** The 6 March issue of *Nature* will include Product Review coverage of this year's Pittcon exhibits plus an article on laboratory automation. The Pittcon theme is analytical chemistry and applied spectroscopy.
- **Microbiology.** The 20 March issue focuses on equipment, services and disposables for the microbiologist.
- **New on the Market.** On 3 April, the latest ideas from manufacturers.
- **FASEB.** In the 10 April issue, news of the exhibition held at St Louis by the Federation of American Societies for Experimental Biology. And more on streptavidin-biotin.
- **Later this year** Product Reviews will be touching on computers, microscopy and other growth areas. □

ADVERTISEMENT

New journal
launched
December 1984

A NEW JOURNAL FROM MACMILLAN PRESS

Pharmaceutical Medicine

Editor Dr R. N. Smith
Assistant Editor P. J. Keen

... "a quarterly journal, which includes papers on the validation of new and established methods of testing drugs in man written by investigators both inside and outside the pharmaceutical industry. It also provides a forum for the exchange of ideas about issues affecting drug development. These new developments merit support from all involved in the management, sale and use of drugs." *Lancet*

Contents of issue one:

EDITORIAL - Pharmaceutical Medicine - Promoting a Discipline. **COMMENTARY** - The Orphaned Clinical Trials: How Many Authors Make a Paper? Balancing Therapeutic Risks and Benefits. New Testament. **REGULATORY SCENE** - N. Mattison Unexpected Side-effects: The Regulatory Implications of the Oraflex and Zomax Withdrawals in the USA. **PAPERS** - G. T. McInnes, J. R. Shelton, I. R. Harrison, J. M. Clarke and R. M. Perkins Assessment of the Long Term Effects of Aldosterone Antagonists by a Short Term Assay. Z. Subhan and I. Hindmarch Assessing Residual Effects of Benzo-diazepines on Short Term Memory. C. Y. Hong, B. N. Chiang and P. Wu In Vitro Spermicidal Potencies of the Dosage Forms and Active Ingredients of Some Vaginal Contraceptives: Measured with a Trans-Membrane Migration Method. R. J. Tallarida, T. M. Smith, A. F. Johnson and R. C. Blodgett Non-Physicians and Physicians Assess Severity of Disease States and Adverse Drug Reactions: Application to Drug Benefit - Risk Measurement. M. J. Vandenburg, K. A. Young, J. Wojtulewski, R. Jenkins, G. Kimber and W. J. C. Currie Does It Matter how Visual Analogue Scales Are Used in Clinical Studies? **BOOK REVIEWS** - Drug Dosage and Administration Strong Medicine. **ABSTRACTS. Instructions for Authors.**

Contents of issue two:

OBITUARY - Derek C. Quantock, PhD, MD. **EDITORIAL** - Therapeutic Trials - Responsibility and Accountability. **COMMENTARY** - Death Certificates and Drug Safety. Dr Derek Quantock - an Appreciation. Dr Graham Dukes - a Chair. Diploma in Pharmaceutical Medicine: Where Do We All Come From? Eyeing Adverse Events. Successful Candidates (Dip. Phar. Med.). **REGULATORY SCENE** - N. Mattison Sources of Change Affecting the US Pharmaceutical Industry in the 1980s. **PAPERS** - S. B. Banes and P. Turner Ophthalmic Opticians and Ocular Adverse Drug Reactions. K. P. Hinchcliffe, K. E. Surrall and J. S. Dixon Reproducibility of Pain Measurements in Rheumatoid Arthritis by Patients using Visual Analogue Scales. A. Johnston, S. Warrington and P. Turner An Attempt to Estimate the Absolute Bioavailability of Flecainide without Reference to a Parenteral Dose. **REVIEW** - B. Whiting and A. W. Kelman The Population Approach to Kinetic Parameter Estimation. **OCCASIONAL LECTURE** - M. N. G. Dukes Pharmaceuticals: Take Another Look. **BOOK REVIEWS** - Guide to Clinical Studies and Developing Protocols. A Manual of Adverse Drug Interactions, 3rd edn. Development of Anti-asthma Drugs. Pharmaceutical Medicine. Whose Body Is It? **LETTERS TO THE EDITOR** - A. E. F. Chandler The Comings and Goings of Medical Advisers. C. Maxwell Orphaned Clinical Trials. **ABSTRACTS. Instructions for Authors.**

Issue three will include:

ABSTRACTS. W. F. McGhan and L. H. Block An analysis of drug recalls and court actions in the US pharmaceutical industry. J. Hill and H. A. Bird Clinical assessments in rheumatoid arthritis. N. Mattison Pharmaceutical innovation and generic drug competition in the US.

For a sample copy of *Pharmaceutical Medicine* issue one please fill in the details below and return to the address shown.

NAME _____

POSITION _____

ADDRESS _____

Return to Alison Baverstock, Macmillan Press,
Houndmills, Basingstoke, Hants RG21 2XS, UK.

Macmillan Press
Scientific & Medical